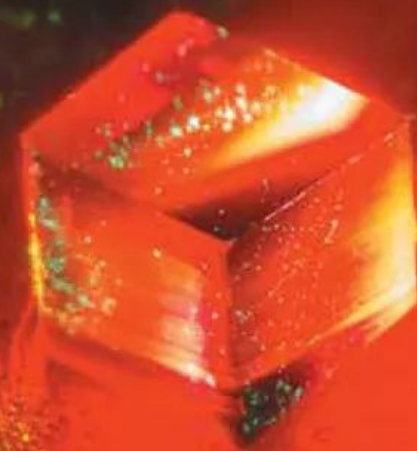


nature

THE INTERNATIONAL WEEKLY JOURNAL OF SCIENCE



THE MASER GOES MAINSTREAM

Diamond defects drive continuous
emission for room-temperature
microwave lasers **PAGES 447 & 493**

STEM CELLS

POTENT FORCE

*Twenty years of research on
human embryonic stem cells*

PAGE 428

SOCIETY

CRISPR CONUNDRUM

*Time to change the
conversation on gene editing*

PAGES 435 & 438

MEDICAL RESEARCH

CANCER CLASSIFIED

*Machine learning tackles
tumour diagnosis*

PAGES 446 & 469

NATURE.COM/NATURE

22 March 2018

Vol. 555, No. 7697

THIS WEEK

EDITORIALS

UNEVEN Asymmetry could offer a path from science to art **p.414**

WORLD VIEW South Korea's public consultation on atomic power **p.415**



FINAL CHANCE Rhino species down to two as last male dies **p.418**

Transformational thinking

The mathematics of Joseph Fourier, born 250 years ago this week, shows the value of intellectual boldness — influencing everything from data processing to machine-learning algorithms.

When you listen to digital music, the harmonies and chords that you hear have probably been reconstructed from a file that stored them as components of different frequencies, broken down by a process known as Fourier analysis. As you listen, the cochlea in your ears repeat the process — separating the sounds into those same sinusoidal components before sending electrical signals to the brain, which puts the components together again.

Fourier analysis allows complex waveforms to be understood and analysed by breaking them down into simpler signals. And it's a shining example of the power and value of intellectual boldness.

The roots of the idea go back to the mid-1700s, when the Italian mathematical physicist Joseph-Louis Lagrange and others studied the vibration of strings and the propagation of sound. But it was one of Lagrange's pupils, Joseph Fourier, who in 1822 truly founded the field that carries his name.

Fourier was born 250 years ago this week, on 21 March 1768. Today, there is virtually no branch of science, technology and engineering that is left untouched by his ideas. Modern versions and analogues of his theory help researchers to analyse their data in almost every discipline, powering everything from YouTube's videos to machine-learning techniques.

Among the scientists who benefited is Ingrid Daubechies, an applied mathematician, who in the 1980s helped to develop the theory of wavelets, which generalized Fourier analysis and opened up previously inaccessible problems. Wavelets were one of the main data-analysis tools used to detect gravitational waves for the first time in 2015, to worldwide acclaim. "He's one of my heroes," Daubechies says.

Before he inspired a revolution in science, Fourier helped to trigger one in his native France. He came of age in the ferment of the 1790s and signed up as a committed *révolutionnaire français* — a decision that almost led to him losing his head to the guillotine during the Reign of Terror that followed the establishment of the First Republic. He joined the army of Napoleon Bonaparte on his invasion of North Africa, alongside dozens of other experts in science, medicine and engineering. With colonial zeal, Napoleon claimed that these intellectuals would help to spread the civilizing values of the Enlightenment.

Fourier worked in Egypt as an administrator, where his efficiency and smart ideas prompted Napoleon to earmark him for a similar position home in France. Back in gloomy northern Europe, Fourier became obsessed with heat and started to apply his mathematical skills to understanding how heat was transferred. He is widely credited as the first scientist to discuss how the greenhouse effect could warm the planet.

He also wanted to understand how heat propagates in a solid object. He discovered the equation that governs this, and showed how to solve it — predicting how the temperature distribution will evolve, starting from the known distribution at an initial time. To do so, he broke the temperature profile down into trigonometric functions, as if it were

a sound wave. Crucially, his analysis included functions for which temperature was allowed to have 'discontinuities', or abrupt jumps. This possibility horrified mathematicians at the time, who were much more comfortable with smooth curves that promised aesthetic simplicity. Fourier stuck to his guns and, as he developed his ideas, started to win his critics over.

Beyond breaking down a function into frequencies, Fourier created a 'dual' profile that encodes all those frequencies, and that became

"Today, there is virtually no branch of science that is left untouched by his ideas."

known as the Fourier transform. In the twentieth century, the Fourier transform became central to quantum mechanics, showing how physical quantities such as position and momentum are 'dual', or complementary, to each other. This means that they cannot be known simultaneously with arbitrary precision: this 'Heisenberg uncertainty' is now seen as one of the fundamental principles of nature. And crystallographers now understand that the X-ray diffraction patterns of a crystal are the Fourier transform of the crystal's structure.

Modern incarnations of Fourier analysis include the 'fast Fourier transform' and 'discrete Fourier transform', which allow faster and more-efficient processing of large amounts of information, including data produced by astronomers.

Fourier would surely be delighted that his ideas have endured. Writing to a friend 229 years ago, he lamented his lack of achievement up to that point: "Yesterday was my 21st birthday; at that age Newton and Pascal had already acquired many claims to immortality."

He succeeded in his fifties. Patience is a virtue, but so is a willingness to pursue intuition to conclusions that conventional wisdom deems illogical. Fourier did that and so stands as a scientific giant who should be remembered and appreciated by researchers everywhere. ■

Getting engaged

Dialogue with the public requires a willingness to accept uncomfortable truths.

Earlier this month, the United Kingdom's Royal Society released the results of a survey of public attitudes to genome-editing technologies. It reported a curious finding: whereas bioethicists like to make the distinction between changes that will and won't be inherited by future generations, the survey respondents didn't. They seemed just as comfortable with genome editing to correct a genetic disorder in embryos as in adult cells. Previous exercises showed the same sentiment.

Public engagement, just like science, can be messy and head in unexpected directions. That cannot, and must not, be a reason not to do it: science has a duty to respond to the views of the public it seeks to serve and represent. And done properly, public engagement can give research more impact and relevance. In general, however, engagement exercises have been viewed by scientists as a one-way transmission of information from experts to the public. This leaves researchers open to the charge that they merely seek public endorsement. Too often, previous attempts to incentivize engagement have burdened individual scientists, who may lack the training, time or funding, resulting in poor-quality engagement and a 'tick-box' mentality.

There are encouraging signs that the scientific community is ready to up its game. Funders such as the US National Institutes of Health and the Wellcome Trust aim to learn more from social-science research on how to improve engagement with stakeholders. The US Food and Drug Administration has established a working group that intends to improve the agency's engagement with patients. The fact that the Royal Society commissioned an analysis of its own efforts, and made the results of that evaluation public, is also good news. However, its failure to seek early input from people who could be first affected by the technology, such as those living with disease or disability, is a missed opportunity.

Change is in the air. In this issue of *Nature*, two Comment pieces suggest ambitious models to improve public engagement (pages 435 and 438), also in the field of genome editing. One calls for a global forum whose members push the discussion beyond the technical abilities of genome editing, and collect a wide diversity of views about its potential applications.

The other article advocates a large consortium that would break down the idea of one, homogeneous 'public' by investigating the distinctions that exist between different communities, such as farmers' unions and parent-and-toddler groups. Smaller engagement processes can also be made more democratic. Researchers could visit participants in their own communities, to encourage open discussion. And when it comes to issues such as genome editing, it is important to include input from a range of stakeholders, such as activists, patient advocates, and church representatives, in the research-planning stages.

"Public engagement, just like science, can be messy and head in unexpected directions."

A World View column this week (page 415) describes an example of how policymakers listened to the public in South Korea, regarding controversial plans to build more nuclear reactors. Faced with growing public concern and even violent protests, the government put together a deliberative poll in which diverse groups of voters were given educational materials and brought together for three days of discussions with experts on both sides of the debate. The resulting poll revealed a surprisingly nuanced stance among the public: ongoing construction of nuclear reactors should continue, the majority said, but the government should pull back from plans to build more. The government followed these suggestions; the violent protests stopped.

As these and other efforts spread and become more sophisticated, engagement can become more about consultation and democracy, and less about the marketing of science. That will benefit researchers and the broader public alike. ■

Asymmetry rules

Singular symposium explores the pervasive presence of symmetry violations.

Those who believe that art and science share common ground (and not everyone does) often point to the concept of symmetry. Science, from fundamental physics to developmental biology, prizes symmetry, and Plato equated it with beauty and harmony. Still, art built on geometric symmetry is rare: even the blockish abstract paintings of Piet Mondrian and the psychedelic art of Bridget Riley have scant use for planes of reflection.

There's a much stronger case to be made for asymmetry as a point of intersection. It's a case bolstered by an intriguing conference held in France last week. The First European Asymmetry Symposium in Nice has a vigorously transdisciplinary programme, which aims to focus and encourage research on asymmetry in systems as diverse as the mouse zygote and market economies, chemical structure, Japanese art and neuroscience.

For all their celebration of symmetry in the laws of nature, physicists conjure more from its breaking. The four fundamental forces are presumed to stem from successive symmetry breakings in the very early Universe, and all the riches of condensed matter and crystallography spill forth from reductions of symmetry. The same is true in biology, in which a progressive elaboration of form is a feature of both evolution and development.

It was the symmetry breaking of a presumed-spherical egg that motivated Alan Turing to develop one of the most fertile models for the emergence of form from uniformity, in 1952. He might not have been right about that aspect of morphogenesis, but his model does explain other types of biological patterning, from animal markings to the ridges of the canine palate. The same ideas reach across disciplines to account for patterning in chemical mixtures and windblown sand.

More surprisingly, perhaps, asymmetry could be fundamental to

aesthetics, too — perhaps reflecting what the art historian Martin Kemp calls a 'structural intuition' that lets us discern vitality in organic form; by contrast, geometric perfection creates a sense of sterility.

"There is no excellent beauty that hath not some strangeness in the proportion," said Francis Bacon (the seventeenth-century philosopher, not the twentieth-century artist, although the latter might well have agreed). Symmetry, once grasped, loses any capacity to surprise us.

The value of asymmetry in molecular science is well established. Louis Pasteur postulated the idea of molecular chirality (to describe molecules whose mirror-image forms cannot be superimposed on each other), and chemistry has pursued the idea almost obsessively since. For synthetic chemists, this particular asymmetry is a maddening challenge: natural molecules are full of chirality, but it's very difficult to produce one form selectively.

As Pasteur concluded, chirality has a central role in life's mysterious origin: the puzzle of why, for example, all chiral amino acids in proteins are of the left-handed variety remains unresolved. (Ditto the right-handedness of nucleotides and DNA's right-handed helix.) Why was symmetry broken, and was it by chance or necessity? Did the fundamental asymmetry of physics — the left-right 'parity' violation by the weak force — play a part in biasing the outcome?

Such questions are always worth revisiting. Whether they will mean much to the economists and linguists at the Nice meeting, say — to whom asymmetry typically means non-reciprocity of inter-agent relations and has nothing to do with spatial structure — remains to be seen.

Arguably, there is more common ground here with physicists studying topology, whether in the connectivity of complex networks or in the handedness of electron band structures of 'topological matter'.

Still, in that multiplicity of meaning lies much of the attraction of asymmetry. When symmetry is broken, choices are made: which forking path to take? Why enter this valley and not that one? Why these laws and not those? Why (it seems) more matter than antimatter? Why is quantum spin 'up' and not 'down'?

Making such choices between alternatives, when neither is obviously preferable, is often a dilemma for artists, too. Perhaps what is truly unifying for artists and scientists is the realization that, of all the many possible worlds, asymmetry makes the actual one unique. ■



Let democracy rule nuclear energy

In South Korea, hundreds of well-informed citizens voted on behalf of their country — a technique that should be used more widely, says Ji-Bum Chung.

News sites have lately carried plenty of stories of how South Korean delegates paved the way towards historic talks about denuclearization of weapons on the Korean peninsula. A meeting between the US and North Korean presidents is set for May at the latest.

Meanwhile, another, less-prominent nuclear conversation — this time about power generation — shows how democratic societies can make complicated decisions involving highly technical issues.

Some two dozen reactors provide about one-third of South Korea's electricity. In 2016, the nation became the world's fifth-largest generator of nuclear energy. Being smaller than other top producers, it has the highest density of nuclear power plants on the planet.

Few would have predicted this 60 years ago. Just after the Korean War, the country was one of the poorest in the world. Then came some of the fastest economic growth observed anywhere. Nuclear energy provided a stable energy source that both enabled this 'compressed development' and became a symbol of it.

Since the construction of the first nuclear power plant in the 1970s, the government, nuclear facilities and the energy industry have often made decisions with little input from civic groups. But the past two decades have seen violent demonstrations against proposed nuclear facilities. Public anxieties were heightened by the 2011 Fukushima meltdown in Japan and a 2016 earthquake in Gyeongju, which hosts six nuclear power plants within 50 kilometres of population centres, such as my workplace Ulsan, with more than one million inhabitants.

In May 2017, Moon Jae-in, who had pledged to decrease the number of nuclear power plants, was elected president. He halted the construction of two power plants at Shin-Kori in which more than US\$1 billion had already been invested. Fierce debate erupted.

Fundamentally, this was a dispute between the nuclear industry and environmentalists — a clash over economic growth versus safety. It was also a political conflict. Although surveys done last year show that public sentiment towards nuclear energy in Korea is generally negative, people who prefer the right-wing Liberty Korea Party generally favour nuclear energy. Age is also a factor. Older citizens, who were lived through the period of compressed development, are more positive about the use of nuclear power plants than are younger ones.

Given the passions that nuclear energy incites, countries around the world have asked citizens to vote on policy. National referendums were used in Sweden (1980), Italy (1987) and Switzerland (1990) to determine whether to maintain or shut down nuclear power plants. But referendums can be blunt instruments: the public is often insufficiently familiar with, or interested in, the scientific, social and economic information necessary for sincere deliberation.

The Moon government took a different tack. It used a 'deliberative

poll' to decide whether to proceed with the additional Shin-Kori plants. This technique was developed by James Fishkin, a professor of communication at Stanford University, California, and has been used to debate topics ranging from public-servant career reform in Brazil to political reform in the United Kingdom. In September 2017, a representative sample of 500 voters was selected on the basis of administrative district, gender and age group, and was sent briefing materials. The next month, the group was brought together for three days of discussions with neutral moderators and pro- and anti-nuclear experts. Participants were briefed on the distribution of earthquake fault lines, the safety features and other technological advances in the planned reactors, and the location of reactors near highly populated areas. Discussions were broadcast throughout the country.

The final vote on 15 October was unambiguous but surprising. Nearly 60% of respondents voted to resume construction. Yet 53.2%

voted to decrease the share of nuclear in the country's energy mix, with 35.5% voting to maintain and 9.7% voting to expand it. It was a nuanced position: respondents thought construction at Shin-Kori should continue for economic reasons; they also thought that nuclear energy should be decreased in the long run for safety reasons. Following the poll, the government resumed work on the two plants at Shin-Kori but cancelled plans to construct six more. There have been no violent protests since.

The deliberative poll faced criticism from both sides. The nuclear industry objected to the poll council's adding a question about the future direction of Korea's energy mix when the original

survey asked only about specific projects. Environmentalists complained that pro-nuclear groups had more resources to make their case, including a strong network of government officials and academics, such as the Korea Atomic Energy Research Institute (one of the largest research institutes in the country). Others thought that the deliberation time was too short and the process for selecting topics was poorly defined.

Although this poll was imperfect, it leaves me optimistic. It is the first attempt in Korea to determine energy policy by broadening and deepening public engagement. Energy is a complicated matter that involves many stakeholders representing multiple beliefs and values. There are lessons here, too, for how nations move forward with other emotive technologies, such as gene editing (see pages 435 and 438) and artificial intelligence.

The decision to consult people to uphold the principle of democracy is a hopeful trend in our chaotic world. ■ SEE EDITORIAL P.413

Ji-Bum Chung is an associate professor at the Ulsan National Institute of Science and Technology in South Korea.
e-mail: learning@unist.ac.kr

THE
DELIBERATIVE
POLL FACED
CRITICISM
FROM
BOTH SIDES.

SEVEN DAYS

The news in brief

AWARDS

Maths prize

Canadian mathematician Robert Langlands won the 2018 Abel Prize on 20 March, for discovering surprising connections between algebra, number theory and analysis. In 1967, Langlands proposed a sort of Rosetta stone for mathematics that allows researchers to translate between different fields. Langlands, now 81 and at the Institute for Advanced Study in Princeton, New Jersey, later found rigorous proofs for parts of this 'dictionary'. Other mathematicians greatly broadened the scope of what has become known as the Langlands program. The annual Abel Prize for scientific work in mathematics, awarded by the Norwegian Academy of Science and Letters, comes with an award of 6 million kroner (US\$ 777,000). See go.nature.com/2gj6can for more.

ASTRONOMY

Pulsars detected

China's flagship radio telescope, the Five-hundred-meter Aperture Spherical Telescope (FAST), has discovered 11 pulsars since it opened in September 2016. The National Astronomical Observatories of China announced the discoveries on 13 March, after observatories in other countries had confirmed the objects. Pulsars are compact, highly magnetized stars that rotate rapidly. The Chinese telescope is the largest in the world and is built in a natural depression among karst mountains in China's southern Guizhou Province. So far it has identified 51 pulsar candidates, of which 11 have been confirmed. The observatory will also be used to map interstellar gas and search for extraterrestrial intelligence.



Last male northern white rhino dies

The northern white rhinoceros (*Ceratotherium simum cottoni*) is on the cusp of extinction after the last male of the subspecies died on 19 March. Just two females are now left. The 45-year-old male, named Sudan (pictured), had age-related muscle, bone and skin ailments that worsened before veterinary surgeons at

the Ol Pejeta Conservancy in Kenya humanely killed him. The last surviving females, Najin and Fatu, are Sudan's daughter and granddaughter, respectively. Conservationists are attempting a plan to save the subspecies using *in vitro* fertilization and, possibly, stem-cell technologies, but numerous hurdles remain.

POLICY

Nicotine proposal

On 15 March, the US Food and Drug Administration (FDA) took a step towards lowering the legal limit of nicotine in cigarettes. The agency announced last year that it was considering the measure as a way to reduce smoking, and it has now released a proposed rule that would cut nicotine in cigarettes to a "minimally addictive or nonaddictive" level. Models co-authored by FDA researchers and published on 15 March predict that by 2060, such a policy shift could prevent 16 million people who otherwise would have become smokers from taking up the habit (B. J. Apelberg *et al.*

N. Engl. J. Med. <http://doi.org/cmmv>; 2018). The proposal is open for public comment until 14 June.

Experimental drugs

On 13 March, the US House of Representatives voted down a controversial bill that would have allowed critically ill people to request experimental medicines without first consulting the US Food and Drug Administration (FDA). The Senate unanimously approved a version of the 'right to try' legislation last year, and President Donald Trump endorsed it during his State of the Union address in January. But some medical groups and patient advocates expressed concerns that the measure would offer false

hope and put patients at risk. They also pointed out that the FDA already has a programme through which physicians can ask the agency to let a patient try an experimental drug — and that the vast majority of those requests are granted.

CLIMATE

Climate lawsuit

An environmental group sued the US Department of State on 13 March to obtain records regarding an overdue report on national efforts to prevent global warming. The United States missed a January deadline to file its Climate Action Report with the United Nations, as required every four years under the 1992 UN Framework

NICHOLE SOBECKI FOR THE WASHINGTON POST VIA GETTY

JACK TAYLOR/GETTY

Convention on Climate Change (UNFCCC). The Center for Biological Diversity in Tucson, Arizona, filed the lawsuit in federal court after the State Department failed to respond to requests concerning the report under the Freedom of Information Act. Although US President Donald Trump has decided to withdraw the country from the Paris climate agreement, the State Department remains legally bound to its UNFCCC commitments.

EVENTS

Spy poisoning

International chemical-weapons specialists arrived in Britain this week to investigate the poisoning of a former Russian spy and his daughter in Salisbury on 4 March (pictured, a forensics tent near where the couple were found). The pair remain critically ill. The UK government says that its defence laboratory in nearby Porton Down found that a military-grade nerve agent of a class developed by Russia was used in the attack. The specialists are from the Organisation for the Prohibition of Chemical Weapons (OPCW) in The Hague; independent verification of the chemical involved is crucial, say experts, and the OPCW is the only body that can do this. In



response to the incident, the government said on 15 March that it will invest £48 million to create a Chemical Weapons Defence Centre at Porton Down. Over the past decade, the lab's funding and staff had been cut as part of government austerity measures.

Autumn bushfires

Wildfires have destroyed dozens of homes in Australia this week, fanned by hot, dry winds that put tracts of the country's southeast on bushfire alert. A fire around the coastal township of Tathra in New South Wales has razed about 70 homes since 18 March. Four separate bushfires in the southwest of the state of Victoria have so far destroyed 18 homes and killed hundreds of beef and dairy cattle. At least one of those fires is thought to have been ignited by lightning strikes,

after a long period of dry weather that has seen some regions experience 40 days without rain.

Aid attacks

Militants ambushed health workers carrying out polio vaccinations in a remote region of northwestern Pakistan on 17 March, killing two people and seriously wounding two others. A faction of the Pakistani Taliban claimed responsibility for the attack, according to media reports. Pakistan is one of only three countries in which polio remains endemic, and the nation has attempted to stamp it out. But distrust of polio workers has risen since it was revealed in 2011 that the US Central Intelligence Agency had planted health workers in the nation to collect intelligence on the fugitive al-Qaeda leader Osama bin Laden.

Strike deal rejected

Strikes at UK universities could resume, after the union that represents academics rejected a deal on pensions. Staff at more than 60 institutions walked out for 14 days over the past month in protest against changes to pension payouts that could cost them thousands of pounds a year in retirement. On 13 March, members of the University and College Union spurned an agreement

made between the union and Universities UK, which represents the academic employers, after 6 days of talks. The union had announced on 8 March that it would schedule a further 14 days of strike action during the student exam period if a deal could not be reached.

BUSINESS

Theranos charges

On 14 March, the US Securities and Exchange Commission (SEC) charged health-technology company Theranos, and its chief executive Elizabeth Holmes, with fraud. Theranos, based in Palo Alto, California, claimed that its blood analysers could perform a range of laboratory tests on just a few drops of blood, according to court documents. The SEC alleges that the company raised more than US\$700 million from investors with false and misleading statements about its analysers and financial performance. As part of a settlement announced on 14 March, Holmes agreed to pay a US\$500,000 fine, and to not serve as a director or officer of a publicly traded company for ten years. "Neither the Company nor Ms. Holmes admitted or denied any wrongdoing," Theranos said in a statement.

PEOPLE

Hawking dies

Stephen Hawking, one of the world's most celebrated scientists, died at his home in Cambridge, UK, on 14 March, aged 76. His death was marked by tributes the world over that celebrated his seminal contributions to cosmology and black-hole research, and his work as a popularizer of science. Hawking had lived with the degenerative condition amyotrophic lateral sclerosis since his twenties. See page 423 for more.

NATURE.COM

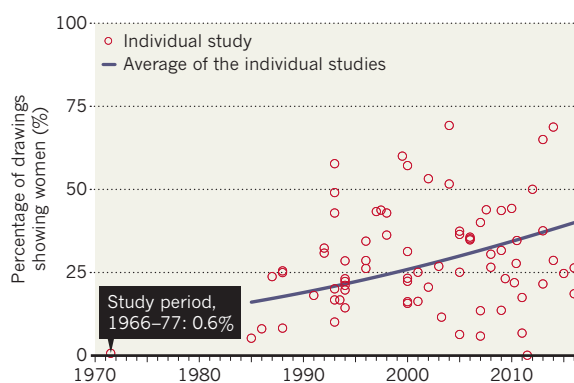
For daily news updates see:
www.nature.com/news

TREND WATCH

Roughly one-third of US children now sketch a woman when asked to draw a scientist — a major shift since the 1960s. In an analysis, psychologists combined the results of 78 'draw-a-scientist' studies between 1966 and 2016, involving some 20,000 US kids from kindergarten to high school. In the 1960s and 1970s, 99.4% of children drew a male scientist; that dropped to an average of 72% between 1985 and 2016. By the 2010s, about one in three drawings portrayed a female scientist.

SKETCHING SCIENTISTS

When asked to draw a scientist, US children today are much more likely than those in the 1970s to doodle a picture of a woman, according to an analysis of five decades' worth of studies.



SOURCE: D. I. MILLER ET AL. CHILD DEV. HTTP://DOI.ORG/CNMR (2018).

NEWS IN FOCUS

COSMOLOGY Stephen Hawking's death prompts outpouring of tributes from scientists **p.423**

ARCHAEOLOGY Excavations in Kenya reveal surprising start to human culture **p.424**

POLITICS China merges its science agencies into one massive ministry **p.425**

PHARMACOLOGY When is an antibiotic too dangerous to prescribe? **p.431**



PIUS UTOMI EKPE/AFP/GETTY



Researchers in Nigeria are racing to identify cases of Lassa fever — and to understand why the ongoing outbreak is the worst on record.

PUBLIC HEALTH

Deadly outbreak tests Nigerian health agency

Reforms made after the Ebola epidemic have boosted Nigeria's capacity to track diseases.

BY AMY MAXMEN

An unprecedented outbreak of a deadly viral disease in Nigeria is showcasing the newfound might of the country's public-health agency. Reforms put in place since a devastating Ebola epidemic struck West Africa in 2014 have transformed how Nigeria responds to infectious diseases — including the current Lassa-virus outbreak.

Since 1 January, Lassa fever has sickened

365 people and killed 81, making it the country's largest recorded outbreak of the virus. But public-health experts say that the toll would be much greater had Nigeria not strengthened its Centre for Disease Control (NCDC) over the past few years. The agency, Nigeria's first line of defence against disease outbreaks, has grown from roughly 30 physicians in 2011 to more than 130 epidemiologists, microbiologists and other specialists today. And it is deploying sophisticated data-management tools and

building diagnostic labs to monitor the current outbreak and prepare for the future.

Later this year, the Nigerian government is expected to approve legislation that would make the NCDC an independent agency with its own budget and decision-making powers.

"The Nigeria CDC has become stronger and faster," says Kingsley Ukwaja, a physician at the Federal Teaching Hospital, Abakaliki in Ebonyi state — a hotspot for the current Lassa outbreak. "They came quickly with protective gear, ►

► and have sent epidemiologists to detect the source of the outbreak, and to locate the contacts of patients who may have the disease.”

Outbreaks of Lassa fever occur regularly in West Africa, where the virus is carried by rats. A person infected with the virus can spread it to others through their blood, urine and other bodily fluids. Symptoms of the disease include fever — and, sometimes, internal bleeding that can lead to death.

The 2014–16 epidemic of Ebola, a virus that can cause similar deadly symptoms, accelerated the transformation of the NCDC. The agency launched in 2011 with medical staff that lacked epidemiological training or the authority to act fast to curb outbreaks, says NCDC chief executive Chikwe Ihekweazu, who works in Abuja.

Nigeria became part of the Ebola epidemic in July 2014, when a man infected with the virus entered the country in Lagos, Nigeria's largest city. Although only 19 people in the country ultimately became infected, many Nigerians felt their country had narrowly skirted disaster. The man in Lagos visited a renowned private clinic where he was diagnosed immediately. That might not have happened had he entered one of the general hospitals that serve much of Nigeria's population, and often lack equipment, doctors and nurses.

In addition, international health organizations were poised to assist Nigerian authorities because they had already been alerted to the

escalating Ebola crisis in Liberia, Sierra Leone and Guinea.

“We were very lucky,” says Ihekweazu, who took the NCDC's helm in 2016 and has changed how it operates. He convinced politicians to grant the agency more autonomy to provide top government officials with information on outbreaks, without concern for the political ramifications. Ihekweazu has also improved how the NCDC interacts with the public. The agency uses television, radio and social media to educate people about public health; that includes combatting false information about how diseases such as Lassa spread.

“We are not where we want to be, but we are miles ahead of where we were,” Ihekweazu says.

DISEASE DETECTIVES

Now, the agency's public-health researchers are trying to understand why the current Lassa outbreak is so extensive. The scientists are collecting samples of the virus to see whether it has become easier to transmit; they are also exploring other hypotheses, such as whether people are living in closer contact to the common African rats (*Mastomys natalensis*) that can carry the virus.

NCDC workers are also building long-term infrastructure and systems to battle this

outbreak and those to come. The agency is coordinating its response at an emergency-operations centre that is modelled on “war rooms” deployed during polio outbreaks and the Ebola epidemic. Staff members are monitoring suspected and potential cases using software called SORMAS, which runs on tablet computers and smartphones. It was developed in response to the plodding pen-and-paper method used to track Ebola.

Yet the NCDC still faces some significant challenges. Many Nigerian states lack facilities to quickly diagnose diseases such as Lassa and Ebola, Ukwaja says, making it harder for the public-health agency to fight outbreaks from the start. In January, three of his colleagues at the hospital in Abakaliki died of Lassa fever. They were exposed to the virus during a four-day wait for test results from samples that had to be shipped out of the state for processing.

Now, that wait has been cut to 24 hours, because the NCDC has built a laboratory in Ebonyi state that has the equipment needed to identify Lassa fever. The facility, which began operating last week, is the fourth such lab in Nigeria.

“Sometimes things need to get worse before they get better,” says Richard Garfield, an epidemiologist who advises the US Centers for Disease Control and Prevention. “Pretty much everything big we have done in epidemiology was in response to a problem we didn't know how to handle.” ■

PUBLISHING

Sites warn against ‘predatory’ journals

Blacklists emerge after closure of popular Beall's list.

BY DALMEET SINGH CHAWLA

When librarian Jeffrey Beall shut down his controversial blog listing potentially ‘predatory’ scholarly publishers and journals last year, copies swiftly appeared elsewhere online. More than a year later, at least one of these copycat blacklists is still growing — maintained by an anonymous website manager who says that they spend hours each weekend working on the list.

Growing interest in the site suggests that there is still an academic appetite for a public blacklist of predatory journals, says the site manager, who identified themselves as a senior research assistant in the hard sciences at a European institution. The site's keeper corresponded with *Nature* by e-mail and declined

to provide any further details of their identity, citing fear of harassment.

Beginning in 2010, Beall, an academic librarian at the University of Colorado Denver, maintained a site listing thousands of open-access journals and publishers that he said deceived authors by charging fees to publish papers without providing expected services, such as peer review and editing. He closed the site in January 2017, and later said that this was because of “intense pressure” from his employer — although his supervisor and institution have denied this, calling it Beall's personal decision.

Sites preserving Beall's list quickly surfaced. The anonymous site manager says that their own page, titled ‘Beall's list of predatory journals and publishers’, was initially intended only for personal use. But soon after the site went live,

they began receiving e-mails from academics asking about the quality of certain journals. The manager now spends four to six hours each weekend replying to these messages, and says that most questions come from academics concerned about publishing in a particular journal.

If the journal titles aren't already listed, the manager says, they carry out an “in-depth analysis” of the publishers' policies, checking them against a set of criteria originally laid out by Beall, and researching whether they are indexed on journal ‘whitelists’, such as the Directory of Open Access Journals or Journal Citation Reports. Journals or publishers deemed untrustworthy by the manager are included in an ‘update’ addendum on the blog. By March 2018, the new site had added 85 stand-alone journals and 27 publishers to Beall's original lists of more than 1,000 titles.

LISTS OF LISTS

Another site, Stop Predatory Journals, also came online in January last year. It is run by a group that, according to the website, consists of scholars and information professionals who decided to “rebuild and resurrect” Beall's list, hoping to create a community-based approach to curation. But the site is not updated regularly.

“I understand their desire to be anonymous,” says Beall, who says that publishers listed

on his site frequently complained to senior officials at his institution.

Last June, a scholarly-services firm called Cabell's International in Beaumont, Texas, launched a pay-to-view blacklist of journals it deems 'deceptive', listing criteria for deciding whether titles should be added. Kathleen Berryman, a project manager at the firm, says that a lack of clear explanations for why journals are on the anonymously maintained blacklists is

a problem. Ideally, every entry on the list would state reasons for its inclusion, agrees the anonymous site manager. "I'm not sure if I will ever have time to do that myself," they say.

Berryman says that around 200 institutions have subscribed to Cabell's blacklist since its launch. The list contains about 8,000 journals, including some that aren't open access. (The firm also maintains a whitelist; some journals aren't on either list, Berryman says.)

Rick Anderson, an associate dean in the library at the University of Utah in Salt Lake City, says that the scholarly community does need a good list of predatory publishers (Anderson did paid consulting work for Cabell's when it was planning its blacklist). But it should include clear criteria and justifications for inclusion, explanations for removed entries and an appeal system, he says. "To do it well is going to be expensive and difficult." ■

PEOPLE

World's scientists pay tribute to Stephen Hawking

The physicist and science icon died at his home in Cambridge, UK, aged 76.

BY DAVIDE CASTELVECCHI

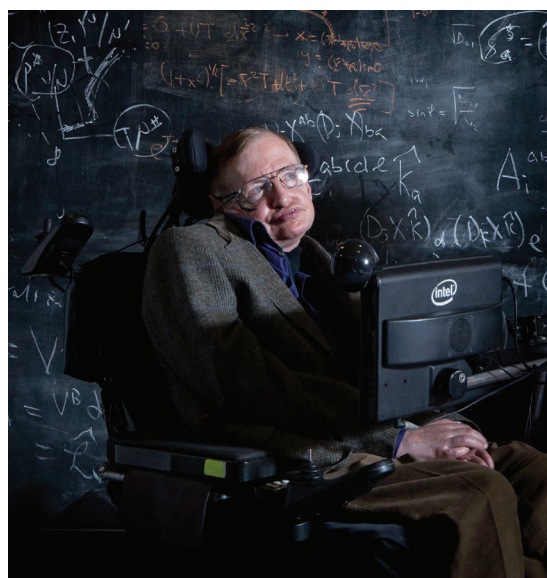
Stephen Hawking, one of the most influential physicists of the twentieth century and perhaps the most celebrated icon of contemporary science, died on 14 March at the age of 76.

Since his early twenties, Hawking had lived with amyotrophic lateral sclerosis (ALS), a disease in which motor neurons die, leaving the brain incapable of controlling muscles. His health had reportedly been deteriorating.

Hawking's death was marked by tributes from scientists worldwide. "The reaction among physicists is just profound shock and sadness," says Malcolm Perry, a theoretical physicist at the University of Cambridge, UK, and a student of Hawking's in the early 1970s. "He was a truly extraordinary man," says Roger Penrose, a theoretical physicist at the University of Oxford, UK, who in 1970 co-authored a seminal paper with Hawking on black holes.

Another former student, theoretical physicist Raphael Bousso at the University of California, Berkeley, told *Nature* that his teacher was a brilliant physicist who also excelled at communicating science to the public. "Stephen was a joyful and light-hearted person, not to be burdened by excessively respectful and convoluted interactions," he says.

The British physicist was born in Oxford in 1942. He was diagnosed with ALS when he was 21, while a doctoral student in cosmology at the University of Cambridge. Physicians gave him just a few years to live, but his disease advanced more slowly than expected. He had an active career for decades, both as a theoretical physicist and as a popularizer of science. Hawking



Stephen Hawking, giant of cosmology, in 2013.

became one of the most recognized names in contemporary science. His books, particularly *A Brief History of Time* (1988), became blockbuster successes. He relished making cameo appearances on television shows such as *Star Trek: The Next Generation* and *The Simpsons*.

Scientifically, his name is most closely associated with the physics of black holes, which he began to study when they were considered mere mathematical curiosities in Albert Einstein's general theory of relativity. In the early 1970s, he began to investigate what quantum physics could reveal about the event horizon, a black hole's surface of no return. Hawking shocked the physics world when he calculated that this surface should slowly emit radiation (soon to become known as Hawking radiation). Black

holes were not truly black.

This emission, he reasoned, should ultimately lead a black hole to shrink and disappear (S. W. Hawking *Nature* **248**, 30–31; 1974). Even more shocking to researchers was Hawking's realization in 1976 that Hawking radiation should erase information from the Universe, in apparent contradiction to some of the basic tenets of quantum theory (S. W. Hawking *Phys. Rev. D* **14**, 2460–2473; 1976). "The importance of this work was not so much the effect itself, but that he was able to provide the one clear-cut physical implication that we know of which brings together the two great revolutions of twentieth-century physics, namely, general relativity and quantum mechanics," says Penrose.

Two years ago, together with Perry and Andrew Strominger at Harvard University in Cambridge, Massachusetts, Hawking began to sketch a possible way out of the black-hole information paradox. The three of them, along with Strominger's student Sasha Haco, had been working on a follow-up paper, which Perry says is in its final stages and will have Hawking's name on it.

Perhaps because most of his work was of a speculative nature and difficult to test, Hawking never won a Nobel prize. In 2016, some wondered whether he might finally win one, when Jeff Steinhauer, a physicist at the Technion–Israel Institute of Technology in Haifa, announced that he had found convincing evidence of Hawking radiation — not in an actual black hole, but in a laboratory analogue made of ultra-cold atoms. However, some experts still consider those results inconclusive.

A more direct test of some of Hawking's findings might yet come from the study of ►

► astrophysical black holes through gravitational waves, initiated by the US-based Laser Interferometer Gravitational-wave Observatory (LIGO). Hawking and others have linked the surface area of a black hole's event horizon to its entropy, a measure of disorder. When interviewed by *Nature's* news team in 2016 about LIGO's first detection of gravitational waves

from merging black holes, Hawking said that he hoped future detections would be sensitive enough to confirm a prediction he had made in the 1970s: that the surface area of a post-merger black hole should exceed the combined surface areas of the original objects that formed it.

Together with cosmologist Thomas Hertog, another of his former students, Hawking had

also explored cosmic inflation — a brief period of rapid expansion in the first moments of the Big Bang — and how it could spawn several universes, a 'multiverse'. The pair set out to transform the idea of a multiverse into a testable scientific framework, says Hertog. "This was Hawking: to boldly go where *Star Trek* fears to tread." ■ [SEE OBITUARY P.444](#)

PALAEOANTHROPOLOGY

Surprise roots for human culture

Technology developments linked to climate turbulence.

BY JEFF TOLLEFSON

Early humans in eastern Africa crafted advanced tools and displayed other complex behaviours tens of thousands of years earlier than previously thought, according to a trio of papers published on 15 March in *Science*¹⁻³. Those advances coincided with — and may have been driven by — major climate and landscape changes.

The latest evidence comes from the Olorgesailie Basin in southern Kenya, where researchers have previously found traces of ancient relatives of modern human as far back as 1.2 million years ago (see 'Complex lives'). Evidence collected at sites in the basin suggests that early humans underwent a series of profound changes at some point before roughly 320,000 years ago. They abandoned

simple hand axes in favour of smaller and more advanced blades made from obsidian and other materials obtained from distant sources. That shift suggests the early people living there had developed a trade network — evidence of growing sophistication in behaviour. The researchers also found gouges on black and red rocks and minerals, which indicate that early Olorgesailie residents used those materials to create pigments and possibly communicate ideas.

A TIME OF CHANGE

All of these changes in human behaviour occurred during an extended period of environmental upheaval, punctuated by strong earthquakes and a shift towards a more variable and arid climate. These changes occurred at the same time as larger animals disappeared from the site and were replaced by smaller

creatures. "It's a one-two punch combining tectonic shifts and climate shifts," says Rick Potts, who led the work as director of the human origins programme at the Smithsonian Institution in Washington DC. "That's the kind of stuff out of which evolution arises."

The studies push back the timeline for such behaviour by around 100,000 years, adding to a growing body of evidence suggesting that the roots of human culture are deeper and more extensive than once thought.

The latest evidence is "probably not enough to put the question to rest as to what effect the climate variability had on human behaviour", says Nick Blegen, an anthropologist at the Max Planck Institute for the Science of Human History in Jena, Germany. But he says that the findings from Olorgesailie provide solid evidence for a shift towards sophisticated behaviour that predates the earliest evidence for *Homo sapiens*. Researchers have traditionally thought that *H. sapiens* emerged around 200,000 years ago, but fossils discovered in Morocco could push that date to more than 300,000 years ago⁴.

Blagen has documented the transport of obsidian in central Kenya roughly 200,000 years ago⁵, and he is preparing another study that would push that record back to 396,000 years ago at the same site. The record for such complex behaviour is likely to extend back even further, he says, but it is not clear whether the environment is shaping human behaviour, or whether advances in human behaviour are enabling them to inhabit riskier environments.

COMPLEX TOOLS

Excavations in the Olorgesailie Basin have been turning up Stone Age artefacts ever since Louis and Mary Leakey pioneered work there in the 1940s. But this is the first time that scientists have documented evidence of more advanced tools and behaviours typically associated with the Middle Stone Age, which lasted until 25,000–50,000 years ago, says Alison Brooks, an anthropologist at George Washington University in Washington DC, who led the dating and analysis of the latest artefacts.

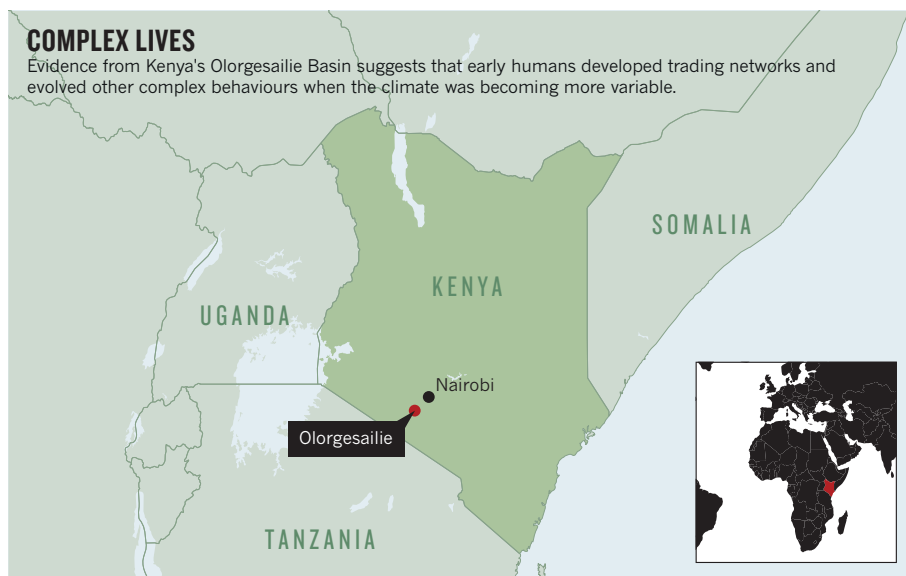
Isotopic dating techniques helped the team to pin down the age of the stone tools, and the researchers traced the obsidian back to its sources, which were mostly located 25–50 kilometres away in multiple directions. "It's the best evidence yet for the exchange of raw materials" so early in time, Brooks says.

Curtis Marean, a palaeoanthropologist at the



Simpler tools (left) gave way to smaller and more complex versions (right) in Kenya's Olorgesailie Basin.

HUMAN ORIGINS PROGRAMME, SMITHSONIAN



projects that drilled into ancient lake beds in Kenya and Ethiopia to collect a detailed record of environmental and ecological changes in the region⁶. Potts and his team drilled two of those cores in the southern Olgosailie Basin, and Potts says the cores cover the entire period that is missing from the archaeological record. Comparisons with cores drilled elsewhere in East Africa should help scientists to differentiate between events happening locally and broader regional climatic trends.

"The drill cores I hope will be a game changer, because of the precision of the environmental record and hopefully the precision of the dating," Potts says. Then it's a matter of working to understand how animals and people might have responded to the changing environment, Potts says. "Only then can we say anything about how climate is really affecting human evolution." ■

1. Potts, R. *et al. Science* <http://dx.doi.org/10.1126/science.aao2200> (2018).
2. Deino, A. L. *et al. Science* <http://dx.doi.org/10.1126/science.aao2216> (2018).
3. Brooks, A. S. *et al. Science* <http://dx.doi.org/10.1126/science.aao2646> (2018).
4. Hublin, J.-J. *et al. Nature* **546**, 289–292 (2017).
5. Blegen, N. J. *Hum. Evol.* **103**, 1–19 (2017).
6. Cohen, A. *et al. Sci. Drill.* **21**, 1–16 (2016).

University of Arizona in Tempe, says he isn't yet convinced by the evidence for trade. "To demonstrate extended social networks, I would like to see regular and systematic transport of raw material across a number of artefact types on the order of 100 kilometres," he says.

The team cannot say exactly how long before 320,000 years these changes happened because an extended period of erosion at the site wiped out the archaeological record there between 499,000 and 320,000 years ago.

Some information could come from several

POLITICS

China creates grand science ministry

Reforms spark fears that fundamental research will suffer.

BY DAVID CYRANOSKI

Chinese leaders released plans last week to expand the powers of the country's science and technology ministry (MOST). The beefed-up agency will continue to oversee science policy and major projects, but will take on extra responsibilities for funding research grants and for recruiting foreign scientists. Politicians say that the reforms will streamline government procedures, but some science-policy experts warn that the changes could weaken support for basic research.

The announcement came at the annual assembly of the National People's Congress in Beijing, where the government revealed that more than 15 ministries and agencies will be merged, restructured or abolished. The National Natural Science Foundation of China (NSFC), the major grant-funding agency, will no longer sit under the powerful State Council, but will be managed by the science ministry.

Other agency changes include expanding

the office that oversees intellectual property, creating a ministry of ecological environment to monitor pollution and forming a conservation agency to help in the protection of endangered species, such as the Przewalski's gazelle (*Procapra przewalskii*).

The plans to expand the science ministry will have the biggest impact on researchers, and took many Chinese scientists by surprise. The NSFC funds modestly sized competitive grants that are initiated by individual investigators. In 2016, its budget of 26.8 billion yuan (US\$3.9 billion) accounted for nearly one-third of China's basic research funding, and was used to support 44,000 research projects. The science ministry, by contrast, manages large projects that are aligned with national goals. Scientists often criticize the ministry for supporting projects on the basis of political and personal connections, rather than expert advice. "Placing NSFC under MOST is likely to complicate these missions," says Cao Cong, a science-policy researcher at the University of Nottingham in Ningbo, China.

Cao, who met with science-ministry colleagues last week, says that many didn't seem to know about the reshuffling. "The reorganization was kept in the dark until the last minute," he says.

Cao says that the reforms could be a sign that the Chinese leadership is unhappy with the progress of the country's previous major reorganization of science, in 2014, which attempted to streamline competitive funding.

The repercussions of the latest reforms will take time to unfold, Cao says. But one scenario is that the NSFC will gradually lose control over basic-research funding, says Cao. "If so, the entire scientific community will be unhappy," he says. NSFC grants are highly regarded by scientists because they are peer-reviewed and place less emphasis on scientists' personal connections. They're also one of the only sources of funding for new PhD students.

Many prominent Chinese researchers say that China's basic-research spending is already too low compared with that of other nations that invest heavily in science. In 2017, China spent 92 billion yuan on basic research, or 0.1% of its gross domestic product (GDP). The United States invests 0.2% of its GDP in basic science.

The NSFC's new head, Li Jinghai, who took over the reins last month, told *Nature* that details about the integration of MOST and the NSFC "need to be discussed and figured out in the coming months". But he said the State Council had promised more money for basic science in a document released in January. "I am sure that basic science in China will be further strengthened," he says. ▶



The endangered Przewalski's gazelle comes under the remit of the new conservation ministry.

► Wang Yifang, director of the Institute of High Energy Physics in Beijing and a member of the National People's Congress, has high hopes for the merger. He has been an outspoken critic of China's process for selecting major scientific projects, saying that peer reviewers sometimes don't have appropriate expertise. Wang wants China to adopt a system more like that in the United States, where expert groups assess the quality of projects.

Conservationists also welcomed plans to create new ministries to protect the environment. The government said an ecology ministry will monitor pollution and enforce environmental-protection laws, and a ministry of natural resources will manage conservation.

Environmentalists say the revamped structure should reduce conflicts between ministries that have similar areas of responsibility. For instance, under the current structure, grasslands

are managed by the agriculture ministry, whereas the wildlife on the grasslands is the responsibility of the State Forestry Administration. As a result, in Qinghai province in central China, the forestry administration has tried to remove fences in grasslands to link up isolated Przewalski's gazelle populations at the same time as the agriculture ministry has funded the construction of new fences. "It's one ecosystem. [It] needs to be addressed as a whole," says Zhang Li, a conservation biologist at Beijing Normal University.

Zhang says that a lack of coordination between agencies has also meant that China's endangered-species list has not been updated for ten years. For instance, the current list designates the Chinese pangolin (*Manis pentadactyla*) — a scaly mammal — and the Yangtze finless porpoise (*Neophocaena asiaeorientalis asiaeorientalis*) as vulnerable to extinction. But the pangolin population has plummeted because of illegal hunting for traditional Chinese medicine and the International Union for Conservation of Nature (IUCN) now lists it as critically endangered. Zhang says that protections for the porpoises, also critically endangered on the IUCN list, should also be strengthened.

Zhang expects that the new natural-resources ministry will update the endangered species list in the near future. ■



The cells that sparked a revolution

After 20 years of hope, promise and controversy, human embryonic stem cells are reshaping biological concepts and even starting to move into the clinic.

BY DAVID CYRANOSKI

Dieter Egli was just about to start graduate school in 1998 when researchers first worked out how to derive human embryonic stem cells. In the two decades since, the prolific cells have been a fixture of his career. The biologist, now at Columbia University in New York City, has used them to explore how DNA from adult cells can be reprogrammed to an embryonic state, and to tackle questions about the development and treatment of diabetes. He has even helped to develop an entirely new form of human embryonic stem cell that could simplify studies on what different human genes do¹.

His wide-ranging research established him as a leader in embryonic

stem-cell biology, a field challenged by restricted funding and an enthusiasm for competing technologies that don't carry the same ethical baggage. Still, many say that human embryonic stem cells are now more relevant than ever. "I am very excited about embryonic stem cells," says Egli. "They will lead to unprecedented discoveries that will transform life. I have no doubt about it."

Embryonic stem (ES) cells provide unparalleled information on early development. Like astronomers looking back to the Big Bang

Neural rosettes, derived from human embryonic stem cells, assemble into spheres in culture.

for fundamental insight about the Universe, biologists rake over the molecules inside these remarkable entities for clues as to how a single original cell turns into trillions, with a dizzying array of forms and functions. Scientists have learnt how to turn the cells into dozens of mature cell types representing various tissues and organs in the body. These are used to test drugs, to model disease and, increasingly, as therapies injected into the body. Starting with an attempt to repair spinal-cord injuries in 2010, there have been more than a dozen clinical trials of cells created from ES cells — to treat Parkinson's disease and diabetes, among other conditions. Early results suggest that some approaches are working: a long-awaited report this week shows improved vision in two people with age-related macular degeneration, a disease that destroys the sharpness of vision².

"In some ways, it's not a surprise, because 20 years ago we expected it," says Egli, "but I'm still surprised that this promise is becoming a reality."

TENTATIVE BEGINNINGS

In 1981, researchers managed to culture stem cells from mouse embryos. They soon recognized the research potential of these intriguing entities, which can both replicate themselves and be nudged into becoming any of the body's 200-plus cell types^{3,4}. But that trick was not easy to accomplish in primates. It took biologist James Thomson, at the University of Wisconsin–Madison 14 years to achieve it in monkeys⁵. Three years later, using donated embryos that had gone unused in fertility treatments, Thomson struck again, creating the world's first human ES-cell line⁶.

The discovery sparked an ethical firestorm. Critics, mostly from religious circles, argued that embryos constitute human beings, and wanted to prevent any research that involved destroying them. In 2001, US President George W. Bush restricted government funding to research on just a few existing ES-cell lines. The decision effectively forced those intent on carrying out the research in the United States to seek private or state funding, and often to create duplicate laboratories — one for ES-cell research and another for work funded by the US federal government. In other countries, including Germany and Italy, the creation of the cells was banned altogether.

Nonetheless, the research went forward where it could. Investigators in Australia, Singapore, Israel, Canada and the United States, among others, soon reported that they had converted embryonic stem cells into neurons, immune cells and beating heart cells⁷.

Researchers also discussed plans to derive stem cells from embryos made by a process called somatic-cell nuclear transfer — the same method used to create cloned animals such as Dolly the sheep — in which the nucleus from an adult donor cell is transferred into a human egg that has had its nucleus removed. The rationale for this 'therapeutic cloning' was to provide a limitless source of dynamic cells with the same DNA as the cell donor. They started talking about studying complex genetic diseases 'in a dish' and replacing failing organs and tissues in the same way that mechanics replace car parts. There were several false starts, notably in 2005, when investigators found that South Korean scientist Woo Suk Hwang had fraudulently claimed to have isolated stem cells in this way. But by 2013, a team led by Shoukhrat Mitalipov, a stem-cell researcher at the Oregon Health and Science University in Portland, finally succeeded⁸.

Throughout the first 15 years, however, much ES-cell research focused on using the cells to understand pluripotency — the amazing ability to become any type of cell. Bit by bit, scientists have been piecing together the molecular pathways that make it possible. "We learned pluripotency from ES cells," says Mitalipov.

Such research contributed to arguably the biggest innovation in regenerative medicine and biological research in the 2000s: the discovery of induced pluripotent stem (iPS) cells. In 2006, stem-cell biologist Shinya Yamanaka at Japan's Kyoto University worked out how to return

adult mouse cells to an embryonic-like state using just four genetic factors⁹. The following year, he and Thomson achieved the same feat in human cells^{10,11}. The process offered, in theory, the same potential pay-off as therapeutic cloning — a limitless supply of pluripotent cells that are genetically matched to a patient — but without the ethical quandaries.

Many predicted that iPS cells would soon displace embryonic stem cells in the research space, but it didn't happen. The number of ES-cell publications grew rapidly after 2006 and has held pace, at about 2,000 per year since 2012. Part of the reason was that ES cells were the gold standard against which researchers could compare iPS cells. And even today, there are some who doubt the safety of using iPS cells. Zhou Qi, a stem-cell biologist at the Chinese Academy of Sciences Institute of Zoology in Beijing, says that concerns that iPS cells would cause tumours inspired him to use ES cells for more than a dozen clinical trials he is orchestrating.

Much of the research on ES cells has been in making them easier to work with. Deriving them was initially a finicky process: plucking one from a culture and growing it into a new population worked less than 1% of the time. A handful of advances has changed that. In 2007, for example, Yoshiki Sasai at the RIKEN Centre for Developmental Biology in Kobe, Japan, discovered a molecule, called a ROCK inhibitor¹², that could keep ES cells from dying when they were removed from the colonies in which they thrived. The success rate for creating new colonies shot to 27%. "It fundamentally changed what you could do," says cell biologist Malin Parmar at Lund University in Sweden. Parmar, who is using ES cells to derive neurons for a Parkinson's disease clinical trial, says that such technical advances ushered in "a new golden age" for ES-cell research.

The cells can now be produced quickly, reliably and indefinitely. And yet they somehow avoid turning into cancer, as some feared they would. "We still do not know why or how" they maintain this equilibrium, says Hiromitsu Nakauchi, a stem-cell biologist at the University of Tokyo, who has been trying to make unlimited supplies of blood platelets from ES and iPS cells.

TIME TO DIVERSIFY

Researchers are also trying to grow organs. Given the right signalling molecules and 3D environment, ES cells organize into complex tissues known as organoids, even in a dish. This capacity is important for researchers such as James Wells at Cincinnati Children's Hospital in Ohio, who is developing intestinal organoids for testing drugs, and perhaps one day for transplant.

And new sources of ES cells have presented other research tools for genetic disease. In 2004, for example, fertility doctors in Chicago started making ES-cell lines from embryos created through *in vitro* fertilization that had been found to have a genetic defect, and thus were rejected for fertility treatments. This allowed the team to create cellular models of thalassaemia, Huntington's disease, Marfan's syndrome, muscular dystrophy and other genetic conditions¹³. In 2007, researchers used ES cells to pin down the molecular changes that lead to cognitive impairments seen in a heritable condition known as Fragile X syndrome¹⁴.

Researchers say that iPS cells promise even more for disease-in-a-dish studies — namely the ability to grow stem cells from any living person with a suspected genetic condition. But many investigators still see strong potential for ES cells in this area. Some conditions cause damage to adult cells that would make any iPS cells derived from them uninformative. And ES cells still have a supporting role.

In 2008, for example, Kevin Eggan at Harvard University in Cambridge, Massachusetts, produced iPS cell lines from people with the neurodegenerative disease amyotrophic lateral sclerosis (ALS). From previous work with ES cells, Eggan knew how to coax pluripotent cells into becoming motor neurons, the brain cells affected by the disease. When he did the

*"In some ways,
it's not a surprise,
because 20 years
ago we expected it."*

same with patient-derived iPS cells, he was able to quickly compare the two types of cell. Cells from patients fired much more than their counterparts from people without the disease¹⁵. “We took advantage of all the work we had done with ES cells to understand motor neurons,” says Eggan. Now, an anti-seizure medicine that quieted iPS cells made from patients is being tested in humans. Results are expected in the next two months.

Egli and Nissim Benvenisty at the Hebrew University of Jerusalem overturned long-held concepts of human biology when they derived ES-cell lines with just half the normal number of chromosomes¹. Researchers are now starting to use gene-editing tools on these ‘haploid’ ES cells to understand how genes function in development. Because they have only one set of genes to worry about, the cells could deliver much more straightforward results, Egli says.

The advances in disease research with ES cells have not all come smoothly. It took Douglas Melton at the Harvard Stem Cell Institute in Cambridge 15 years to turn ES cells into functional β -cells — the pancreatic cells that can sense glucose and produce insulin. Then he couldn’t find any difference between pancreatic cells produced from normal ES cells and iPS cells from people with type 1 or 2 diabetes. “We know there is a genetic susceptibility, but that doesn’t mean you can see it *in vitro*,” he says.

CELL REVIVAL

Melton still has plans for the β -cells he’s made from ES cells. He hopes to transplant them into people with type 1 diabetes to end, or at least reduce, their reliance on insulin injections. The last hurdle in the work is introducing the cells so that they are not destroyed by the immune system. Semma Therapeutics, a company that Melton founded in Cambridge, aims to do this by ensconcing the cells in a pouch that would allow nutrients in and insulin out, but would block access to immune cells. He expects to start clinical trials within three years. ViaCyte in San Diego, California, has just restarted a similar clinical trial it launched in 2014 after redesigning its encapsulation technology. And other companies, such as Novo Nordisk in Denmark are starting up programmes for diabetes using cells derived from ES cells.

In the clinical realm, many have assumed that iPS cells would eventually win out over ES cells. One potential advantage is that they can produce cells and tissues with the same DNA as the patient and thus not cause an immune reaction when transplanted. But for most genetic diseases, including type 1 diabetes, iPS cells created from a patient would contain the mutation that causes the problem, and the cells would have to be modified to confer any therapeutic benefit.

Then there’s the matter of cost. Preparing a single iPS-cell line for clinical use would cost roughly US\$1 million, says Jeanne Loring, a stem-cell biologist at the Scripps Research Institute in La Jolla, California. That’s currently prohibitive if the goal is to use a patient’s own cells, but Loring expects that the price will come down and is working on developing iPS cells as a treatment for Parkinson’s disease.

So far, researchers have initiated just one human trial using cells derived from iPS cells. Led by ophthalmologist Masayo Takahashi at the RIKEN Center for Developmental Biology, it aims to treat macular degeneration, but was halted in 2014 when investigators decided to simplify the procedure and use donor-derived, rather than patient-derived, stem cells. It restarted in 2017, but hit another roadblock in January, when a membrane developed in the eye of a participant and had to be surgically removed.

Macular degeneration has been a popular target for ES-cell therapies. There have been at least six clinical trials, in the United States, the United Kingdom, South Korea, China and Israel. On 19 March, researchers led by ophthalmologist Pete Coffey, director of the London Project to Cure Blindness and the University of California, Santa Barbara, reported the results of a study to implant a patch of cells made from ES cells into the damaged retinas of two individuals². A year after the procedure, the participants regained the ability to read, albeit slowly.

“We took advantage of all the work we had done with ES cells to understand motor neurons.”

Alan Marmorstein, an ophthalmologist at Mayo Clinic in Rochester, Minnesota, calls it a “big step forward” for the field. “This is the first strong indication of efficacy in humans and it certainly supports further studies in other parts of the body,” he says. Coffey says the breakthroughs are finally arriving because scientists are now working out how best to put the cells into people. “A decade ago, we thought, ‘You just needed to put the cells in, and the cells will know what to do.’ That’s not true — they have to be controlled in some appropriate way.” Many in the stem-cell field are betting the next big clinical breakthrough for ES cells will come in Parkinson’s disease. The disorder is caused by a loss of the neurotransmitter dopamine, and half a dozen companies and clinics are gearing up to use ES cells or iPS cells to replace dopamine-producing neurons.

One crucial question is how far the pluripotent cells should be taken down the road towards maturity before transplanting them. An Australian trial started in 2016 and a Chinese trial begun in 2017 use immature neural precursor cells, which do not produce dopamine. The researchers say the immaturity of the cells will help them to survive transplantation and integrate into their new host’s brain. But leaders of a group of ES- and iPS-cell trials known collectively as GForce-PD say that the more-mature cells they use turn into the desired type of dopamine-producing cell more reliably and are less likely to grow out of control.

PATHWAY TO PROMISE

ES-cell research still has room to grow, if it can get past some hurdles. One big problem is that many cell types are challenging to produce. Melton estimates that only about

ten cell types created so far are truly functional equivalents of normal human cells. And some with the most far-reaching uses, such as eggs and sperm, are expected to remain a challenge for the foreseeable future.

The field also faces uncertainty about funding. Scientists have heard frequent rumours that US president Donald Trump might impose new restrictions on federal funding for research on ES cells.

But despite their sometimes rocky history, ES cells have proved their value repeatedly, and in some unpredictable ways, say many investigators. Some researchers have even scaled back their use of animal models because ES cells seem to provide a better path to studying human disease. “My motto was, ‘all human, all the time,’” Melton says.

Yamanaka says that ES cells were the motivation for his own work on iPS cells. And it was Thomson’s recipe for human ES cells that allowed the shift from mouse to human iPS cells in just one year, after it had taken nearly two decades to move from mouse ES cells to the human variety. “We knew exactly how we should culture human iPS cells,” says Yamanaka.

ES cells are just as crucial today, he says, for better understanding the mechanism of pluripotency and for improving the medical application of any pluripotent cell. “The importance of human ES cells is no less now than 20 years ago, and I do not imagine it will be any lower in the future,” he says. ■

David Cyranoski writes for Nature from Shanghai, China.

1. Sagi, I. *et al. Nature* **532**, 107–111 (2016).
2. da Cruz, L. *et al. Nature Biotechnol.* <http://dx.doi.org/10.1038/nbt.4114> (2018).
3. Martin, G. R. *Proc. Natl Acad. Sci. USA* **78**, 7634–7638 (1981).
4. Evans, M. J. & Kaufman, M. H. *Nature* **292**, 154–156 (1981).
5. Thomson, J. A. *et al. Proc. Natl Acad. Sci. USA* **92**, 7844–7848 (1995).
6. Thomson, J. A. *et al. Science* **282**, 1145–1147 (1998).
7. Murray, C. E. & Keller, G. *Cell* **132**, 661–680 (2008).
8. Tachibana, M. *et al. Cell* **153**, 1228–1238 (2013).
9. Takahashi, K. & Yamanaka, S. *Cell* **126**, 663–676 (2006).
10. Takahashi, K. *et al. Cell* **131**, 861–872 (2007).
11. Yu, J. *et al. Science* **318**, 1917–1920 (2007).
12. Watanabe, K. *et al. Nature Biotechnol.* **25**, 681–686 (2007).
13. Verlinky, Y. *et al. Reprod. Biomed. Online* **10**, 105–110 (2005).
14. Eiges, R. *et al. Cell Stem Cell* **1**, 568–577 (2007).
15. Dimos, J. T. *et al. Science* **321**, 1218–1221 (2008).

When antibiotics turn *toxic*

Researchers are asking why fluoroquinolone drugs cause rare, disabling side effects.



BY JO MARCHANT

In 2014, Miriam van Staveren went on holiday to the Canary Islands and caught an infection. Her ear and sinuses throbbed, so she went to see the resort doctor, who prescribed a six-day course of the popular antibiotic levofloxacin. Three weeks later, after she had returned home to Amsterdam, her Achilles tendons started to hurt, then her knees and shoulders. She developed shooting pains in her legs and feet, as well as fatigue and depression. “I got sicker and sicker,” she says. “I was in pain all day.” Previously an active tennis player and hiker, the 61-year-old physician could barely walk, and had to climb the stairs on all fours.

Since then, she has seen a variety of medical specialists. Some dismissed her symptoms as psychosomatic. Others suggested diagnoses of fibromyalgia or chronic fatigue syndrome. Van Staveren is in no doubt, however. She’s convinced that the antibiotic poisoned her.

She’s not alone. Levofloxacin is one of a class of drugs called fluoroquinolones, some of the world’s most commonly prescribed antibiotics. In the United States in 2015, doctors doled out 32 million

prescriptions for the drugs, making them the country’s fourth-most popular class of antibiotic. But for a small percentage of people, fluoroquinolones have developed a bad reputation. On websites and Facebook groups with names such as Floxie Hope and My Quin Story, thousands of people who have fallen ill after fluoroquinolone treatment gather to share experiences. Many of them describe a devastating and progressive condition, encompassing symptoms ranging from psychiatric and sensory disturbances to problems with muscles, tendons and nerves that continue after people have stopped taking the drugs. They call it being ‘floxed’.

For decades, regulatory agencies and the medical profession were sceptical that a brief course of antibiotics could have such a devastating, long-term impact. But after persistent campaigning by patient groups, attitudes began to change in 2008, when the US Food and Drug Administration (FDA) announced the first of what would be a series of strong alerts about the side effects of fluoroquinolone drugs, including tendon

ILLUSTRATION BY OLIVER MUNDAY

rupture and irreversible nerve damage. In 2016, the agency accepted the existence of a potentially permanent syndrome that it calls fluoroquinolone-associated disability (FQAD), and recommended that the drugs be reserved for serious infections. That move has triggered other regulatory agencies to reassess the antibiotics: Health Canada warned doctors of rare cases of persistent or disabling side effects in January 2017, and the European Medicines Agency (EMA) is expected to publish the results of a safety review this year, after a public hearing planned for June.

Fluoroquinolones are valuable antibiotics, and safe for most people. Yet they are so widely prescribed that their side effects might have harmed hundreds of thousands of people in the United States alone, say scientists who are working with patients to unpick FQAD's causes. Fluoroquinolone toxicity, they say, provides a compelling example of an emerging understanding that antibiotics don't just harm microbes — they can severely damage human cells, too. Until recently, investigations into the side effects of antibiotics have focused on how the drugs disrupt the human microbiome, says James Collins, a medical engineer at the Massachusetts Institute of Technology in Cambridge. "Antibiotics are also disrupting our cells, and in pretty hefty ways," he says.

THE DARK SIDE OF FLUOROQUINOLONES

Quinolone antibiotics, first developed in the 1960s, kill bacteria by blocking enzymes called class II topoisomerases, which normally untangle DNA during cell replication. These enzymes usually cut DNA's double helix, pass another part of the strand through the gap, and then mend the cut. But quinolones bind to the enzymes, preventing them from mending their cuts. In the 1980s, researchers added fluorine atoms to the quinolones' structures. This allowed the antibiotics to penetrate tissues throughout the body, including the central nervous system, and boosted their effectiveness against a broad range of bacterial infections.

Some FDA-approved fluoroquinolones were swiftly withdrawn from the market after severe adverse reactions and several deaths — trovafloxacin, withdrawn in 1999, damaged livers, for instance. But others became the drug of choice both for serious infections and for routine complaints, despite rare side effects. "These are heavily used drugs because they are very effective," says Joe Deweese, a biochemist who studies topoisomerases at Lipscomb University College of Pharmacy in Nashville, Tennessee. In the 1990s, ciprofloxacin (cipro) was given to US troops serving in the Persian Gulf as prophylaxis in case of exposure to anthrax spores. And in 2001, sales of cipro surged after a series of terrorist attacks involving anthrax; the US Centers for Disease Control and Prevention (CDC) recommended a 60-day course for anyone at risk of being exposed.

But by that point, some people had already flagged potential problems. In 1998, US journalist Stephen Fried (now at Columbia Journalism School in New York) published a book called *Bitter Pills* about his wife's severe and long-lasting neurological reaction to ofloxacin. It helped to trigger a wave of reports on websites such as the Quinolone Antibiotics Adverse Reaction Forum, which by 2001 hosted more than 5,000 posts. The late Jay Cohen, then a psychiatrist and medical researcher at the University of California, San Diego, contacted patients through the sites and published 45 case studies¹. Cohen warned that after taking fluoroquinolones, some people had developed serious problems in multiple organs. These effects came on rapidly and lasted for months or years.

Cohen's work was largely dismissed at the time because of his reliance on online forums. But complaints and patient petitions continued. From the 1980s to the end of 2015, the FDA received reports from more than 60,000 patients detailing hundreds of thousands of 'serious adverse events' associated with the 5 fluoroquinolones still on the market (most commonly tendon rupture, as well as neurological and psychiatric symptoms), including 6,575 reports of deaths. The FDA says that the reports of adverse events it receives — sent in by drug manufacturers, by doctors

and directly by consumers — cannot be used to reach conclusions about the severity of problems associated with drugs. Still, the fluoroquinolones have attracted more complaints than other more widely used antibiotics. And only 1–10% of adverse events are estimated to be reported to the FDA, suggesting that fluoroquinolones might have harmed hundreds of thousands of people in the United States alone, says Charles Bennett, a haematologist at the University of South Carolina's College of Pharmacy in Columbia. Bennett is also director of the Southern Network on Adverse Reactions, a state-funded pharmaceutical-safety watchdog, which has been working with people affected by fluoroquinolones since 2010.

In 2008, the FDA announced 'black box' warnings of tendon rupture among those given the antibiotics; in 2013, it added a risk of irreversible nerve damage. (Such warnings are placed inside a black box on drug labels, and call attention to serious or life-threatening risks.) As alerts

mounted, patients launched lawsuits against manufacturers of the drugs, claiming they had not been adequately informed of risks. These cases have been variously won, lost or settled for undisclosed sums, and many are still in progress; manufacturers argue that they handled risks appropriately, and work with the FDA to update safety labels.

In November 2015, the FDA voted to recognize FQAD as a syndrome on the basis of 178 cases that the agency regarded as clear-cut: otherwise healthy people who took fluoroquinolones for minor ailments and then developed disabling and potentially irreversible

conditions². The FDA also noted a disturbing pattern: fluoroquinolones had a much higher percentage of disabilities among their serious-adverse-event reports than did other antibiotics.

MITOCHONDRIAL DAMAGE

Beatrice Golomb at the University of California, San Diego, has been working for a decade with people affected by fluoroquinolones, beginning with David Melvin, a police officer and keen cyclist who had to use a wheelchair after he was given levofloxacin for suspected epididymitis in 2007. Accumulating evidence, Golomb says, suggests that fluoroquinolones are damaging mitochondria, the power packs inside human cells that evolved from symbiotic, bacteria-like cells billions of years ago. This kind of harm can affect every cell in the body, explaining why a wide range of symptoms can appear and get worse over time.

Mitochondrial toxicity is a problem with many classes of drug, says Mike Murphy, who studies the biology of mitochondria at the University of Cambridge, UK. But because mitochondria retain some similarities to their bacterial ancestors, antibiotics can pose a particular threat to them. Researchers have shown, for example, that aminoglycoside antibiotics can cause deafness by damaging mitochondria in the hair cells of the ear³.

Isolated studies from the 1980s onwards have suggested that fluoroquinolones impair mitochondrial function, but a 2013 study⁴ by Collins and his colleagues is the most convincing, researchers say. They reported that antibiotics in several classes triggered oxidative stress — a build-up of reactive, oxygen-containing molecules — in mitochondria, inhibiting their function across a range of mammalian cells, as well as in mice. "We were surprised at how strong the effect was and how common the effect was across the different classes," Collins says. But "the largest effects were seen in the quinolones".

Pharmaceutical researchers had spotted the issue, too: in 2010, toxicologist Yvonne Will and her colleagues at Pfizer in Groton, Connecticut, reported an assay to detect mitochondrial damage early in drug development⁵. They found that some antibiotics affected mitochondria and others didn't. Every fluoroquinolone they tested damaged mitochondria in human liver cells — having what the researchers described as "a strong effect" at therapeutic concentrations, although Will cautions that it isn't possible to extrapolate from that result to clinical outcomes.

But the potential for mitochondrial damage still isn't widely appreciated

"It is almost impossible for doctors to keep up."

among antibiotics researchers and the medical community, Collins says. “I think people generally assume that antibiotics do not impact mammalian cells,” he says. One problem is that there is still no reliable biomarker that researchers can use to test for mitochondrial damage in people, tying cell-line research to clinical experience. Nor is it known precisely how the fluoroquinolones are damaging human cells. A 2013 FDA review of the antibiotics’ safety, for example, cited a 1996 study⁶ reporting that cipro caused DNA breaks in mitochondria in a variety of mammalian cell lines. But Neil Osheroff, a biochemist at Vanderbilt University in Nashville, Tennessee, who studies fluoroquinolones, is doubtful about that result. He has done his own lab tests, and found that, at therapeutic concentrations, the fluoroquinolones prescribed by doctors have very little effect on human DNA⁷. Meanwhile, mitochondrial damage isn’t the only theory in play: a 2015 study, done on human kidney cells⁸, reported that fluoroquinolones can bind to iron atoms from the active sites of several enzymes that modify DNA, leading to epigenetic changes that might be related to some of the drugs’ side effects.

At a conference last September, Bennett reported preliminary data that might hint at why only some people develop serious side effects from fluoroquinolones. He took saliva samples from 24 people who reported neuropsychiatric side effects — such as memory loss, panic attacks and depression — and found that 13 of them (57%) shared a gene variant usually seen in only 9% of the population. Bennett is not revealing the gene’s identity because he has a patent application in process, but he says that it seems to be a site related to poor metabolism of the quinolones. Such a mutation might cause dangerously high levels of the drug to accumulate in cells, including in the brain. Bennett is now conducting a trial with 100 more participants to see if he can replicate the result. If so, that might lead to a genetic test to identify people who should not be given the drugs.

LACK OF SUPPORT

Most scientists asked by *Nature* about fluoroquinolones said that more research is needed to understand their side effects. Collins hopes to explore mitochondrial damage by antibiotics in other animal models. He and Murphy have also found, in lab studies^{4,9}, that giving antioxidants alongside fluoroquinolones seems to mitigate the effects on mitochondria. Murphy is interested in trials to avoid mitochondrial toxicity in drugs; he owns shares in a company that aims to set some up. But such trials are difficult and expensive, particularly for drugs that are given in sometimes life-threatening situations, he says. Golomb is currently conducting an unfunded online survey to gather information on the experiences of thousands of patients. She hopes that it will lead to hypotheses about what might mitigate harms that could then be tested in clinical trials. But little support is available. That’s typical for research on drug safety. Investigating medications that have been on the market for years isn’t a priority for research agencies such as the US National Institutes of Health, says Bennett. Manufacturers don’t have an incentive to fund post-market safety

studies, particularly for off-patent drugs such as cipro and levofloxacin, where the vast majority of sales are from generics firms. “So there is really nobody to champion this work,” says Bennett.

Another factor is scientists’ reluctance to publish results that drug companies might find unfavourable. “There’s a long history of adverse action against people who expose drug and chemical harms,” says Golomb. She

cites a list made by the pharmaceutical firm Merck of doctors who criticized the anti-inflammatory drug Vioxx (rofecoxib), which was withdrawn from the market over an increased risk of heart attack and stroke. According to internal e-mails read out in court in 2009 as part of a Vioxx class-action case in Australia, a list e-mailed among Merck employees contained doctors’ names with the labels “neutralize”, “neutralized” or “discredit” next to them. (Merck did not respond to *Nature*’s request for comment.) Aggressive tactics are “a very big problem”, says Bennett, who says he has been threatened by drug companies in the past.

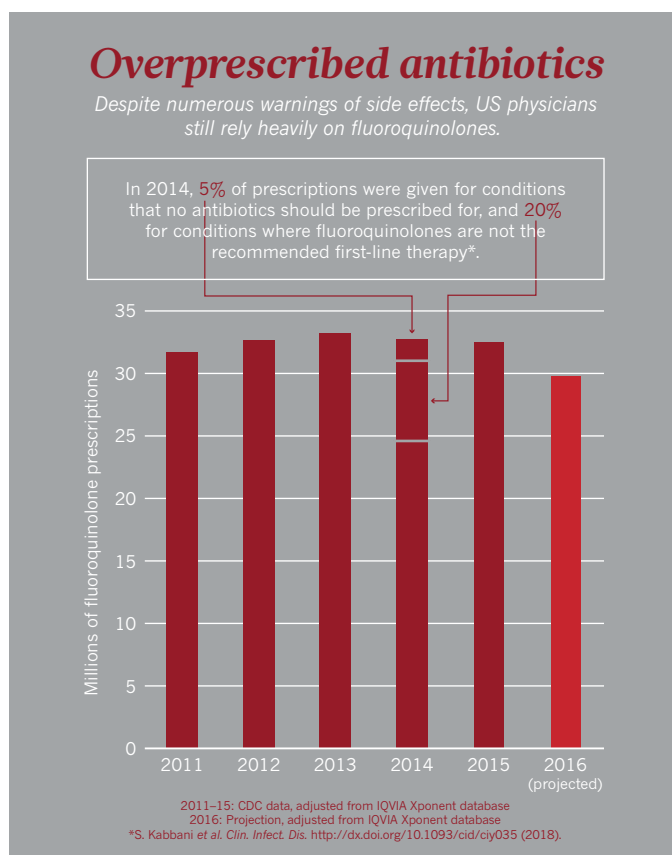
For Osheroff, there are more pressing concerns about fluoroquinolones, such as how to combat the emergence of antibiotic-resistant infections and develop new antibiotics. But he and other scientists agree that doctors should not prescribe fluoroquinolones for relatively minor infections when they could use other drugs. The FDA’s warnings on drug labels, how-

ever, have been slow to produce results. Prescriptions for the drugs did not fall between 2011 and 2015, according to the CDC (see ‘Overprescribed antibiotics’). This suggests, says Bennett, that official alerts haven’t been enough to get physicians to change their habits. The labels for fluoroquinolones alone have changed around 20 times in the past 4 or 5 years, he says. “It is almost impossible for doctors to keep up.” Still, US prescriptions of fluoroquinolone antibiotics did drop by around 10% in 2016, and totals for the first half of 2017 suggest that prescriptions fell again last year, according to unpublished figures given to *Nature* by IQVIA, a health-data firm in Durham, North Carolina.

Meanwhile, van Staveren, in Amsterdam, complains that despite her medical training, she still can’t find a doctor who believes her. She is waiting for the EMA to make its recommendations, and hopes that it will follow the FDA in acknowledging FQAD, and warning about it. (The EMA declined to comment while the review is ongoing.) “I want doctors to be informed about the risks, no matter how rare or not they are,” van Staveren says. “I want warnings all over and I want the warnings to be taken seriously.” ■

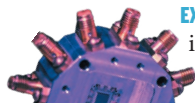
Jo Marchant is a science journalist based in London.

1. Cohen, J. S. *Ann. Pharmacother.* **35**, 1540–1547 (2001).
2. Tennyson, L. E. & Averch, T. D. *Urology Pract.* **4**, 383–387 (2016).
3. Gao, Z., Chen, Y. & Guan, M.-X. *J. Otol.* **12**, 1–8 (2017).
4. Kalghatgi, S. et al. *Sci. Transl. Med.* **5**, 192ra85 (2013).
5. Nadanaciva, S. et al. *Biomol. Screen.* **15**, 937–948 (2010).
6. Lawrence, J. W. et al. *Mol. Pharmacol.* **50**, 1178–1188 (1996).
7. Aldred, K. J. et al. *ACS Chem. Biol.* **8**, 2660–2668 (2013).
8. Badal, S. et al. *J. Biol. Chem.* **290**, 22287–22297 (2015).
9. Lowes, D. A. et al. *Free Radical Res.* **43**, 323–328 (2009).



COMMENT

PHYSICS A clear guide to the disorientating quantum world **p.440**

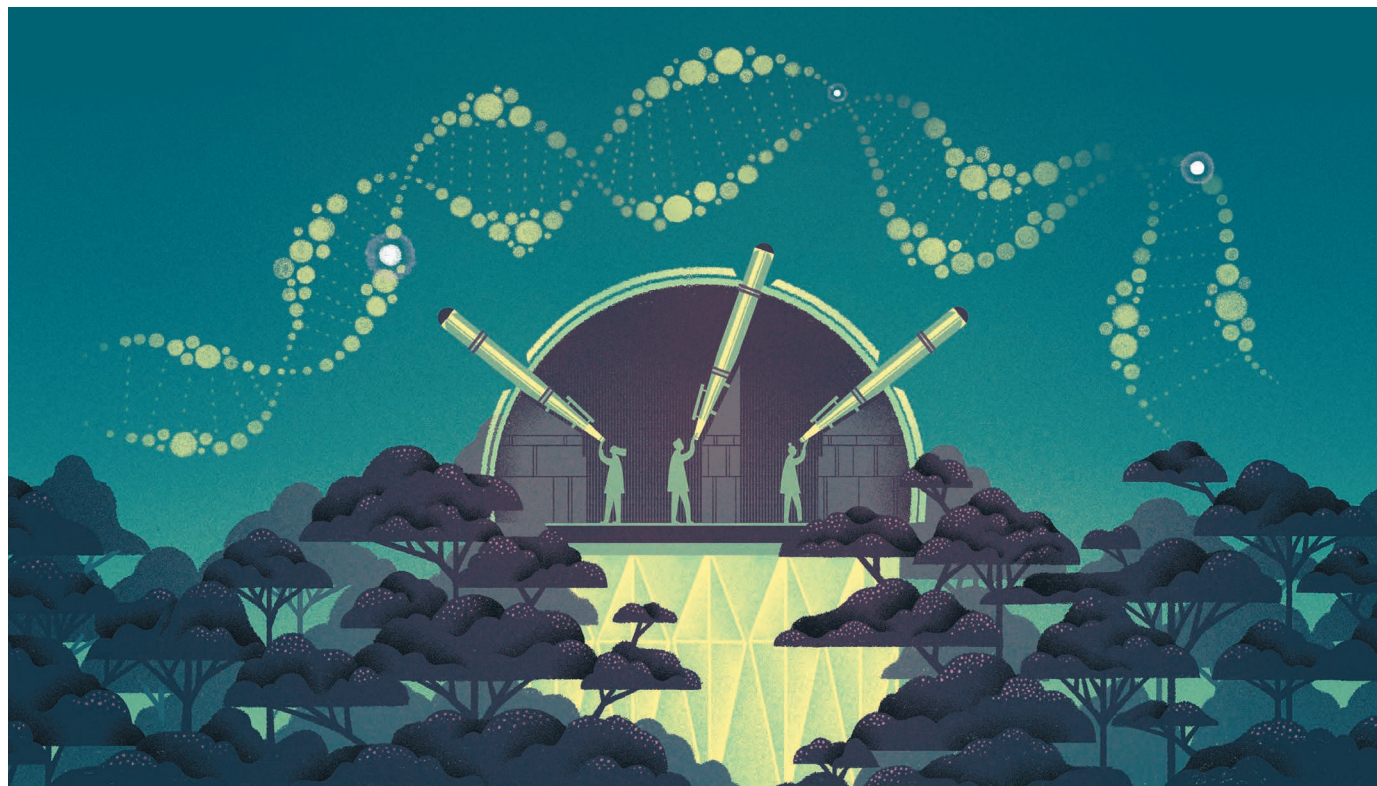


EXHIBITION Good fakes, instructive fictions and phoney baloney **p.442**

EDUCATION Teach digital intelligence to build kids' resilience to online risks **p.443**

OBITUARY Stephen Hawking, iconic astrophysicist, remembered **p.444**

ILLUSTRATION BY MARINA MUJUN



A global observatory for gene editing

Sheila Jasanoff and J. Benjamin Hurlbut call for an international network of scholars and organizations to support a new kind of conversation.

In August 2017, scientists reported that they had used the gene-editing tool CRISPR–Cas9 to correct a mutation in viable human embryos. The work is just one of countless applications of the technique, with which scientists hope to alter plants, animals and humans.

The value of most applications of the technology has barely been exposed to public review. Unless these editorial aspirations are more inclusively debated, well-intentioned research could move humanity closer to a future it has not

assented to and might not want.

Over the past three years, leading scientists have called for global deliberation on the possible effects of gene editing on the human future¹. In our view, the discussions that have taken place fall far short of the expansive, cosmopolitan conversation that is needed.

DOWN FROM THE SUMMIT

An important milestone was the International Summit on Human Gene Editing, held in Washington DC in December 2015.

Organizers called for an international forum to seek “broad societal consensus” on the norms that should guide research².

Nobel laureate David Baltimore began the summit by invoking the 1975 Asilomar meeting on recombinant DNA research³: “In 1975, as today, we believed it was prudent to consider the implications of a remarkable achievement in science. And then, as now, we recognized we had a responsibility to include a broad community in our discussions.”

Asilomar is often remembered as a model of successful self-regulation that affirmed ►

► science's autonomy and the principle of responsible research. Yet at the 2015 summit, as at Asilomar, the questions asked, the forms of expertise called upon, and the definition of stakes for science and human life were all shaped by those communities most aggressively advancing the research.

The summit brought together a more diverse and international group than is typical of meetings on the implications of scientific research. But the discussion still focused on predictions about what genome editing will be able to do in the near term and what its biological risks are, even though it raises issues that clearly transcend immediate concerns for health and safety. Moreover, the meeting format offered little opportunity for deeper listening or learning.

Instead, it encouraged an all-too-common pattern⁴. Discussion split into two camps: scientific experts explored technical issues, whereas scholars who study science and society addressed questions about the possible disruption to social norms. The two camps did not inform each other.

To break out of this bifurcation between the 'science' and the 'ethics', methods must be found to get people to engage substantively with each other. In our view, an entirely new type of infrastructure is needed to promote a richer, more complex conversation — one that does not originate from scientific research agendas but that instead invites multiple viewpoints.

We advocate the establishment of a global observatory for gene editing, as a crucial step to determining how the potential of science can be better steered by the values and priorities of society. This would be an international network of scholars and organizations similar to those established for human rights and climate change. The network would be dedicated to gathering information from dispersed sources, bringing to the fore perspectives that are often overlooked, and promoting exchange across disciplinary and cultural divides.

ALTERNATIVE VISION

In seeking new models, it is worth recalling a little-known meeting held at Airlie House in Warrenton, Virginia in April 1976 — a counterpoint to Asilomar's narrow, expert-dominated approach. There, about 50 participants debated whether a new social contract was needed between society and science⁵. Half of the group were scientists; the rest were lawyers, public-interest advocates, philosophers, journalists and congressional staff members.

At that gathering, the philosopher Stephen Toulmin declared that science was facing the equivalent of the Protestant Reformation that splintered Europe 500 years ago. "People are tired of being shut out of science's ecclesiastical courts and are demanding to be let in," he said. Hans Jonas, another philosopher,

was more blunt. "Scientific inquiry," he said, "demands untrammelled freedom for itself."⁶

Today, a reformation of the contract between science and society is even more overdue, but the institutional barriers are even more entrenched. Certainly in relation to gene editing, there has been much more advice from experts than acknowledgement of the limits of such expertise.

In April 2017, we gathered three dozen social scientists, ethicists, religious thinkers, legal scholars, scientists and representatives of national and international ethics bodies at Harvard University in Cambridge, Massachusetts. We discussed how to enable a different kind of conversation about the variety of techniques with which scientists can edit living systems.

We did not start with the usual question of what science is ready to achieve. Instead, we took a step back and asked to what extent existing scientific and political institutions are capable of initiating the forms of deliberation demanded by the prospect of editing life. We explored the rights and responsibilities of scientific experts, policymakers, publics and scholars in such processes. And we asked what is needed — in terms of representation and deliberation — for a genuinely broad societal consensus on gene editing.

We agreed on the need for a coordinated international effort to gather and analyse salient information on what is already being done to integrate perspectives from science and society. That effort would bring to light divergent ideas about what is at stake in protecting the integrity of life, human and non-human, against unwarranted intrusion from new and emerging technologies.

We identified the need for a forum to promote sustained international, interdisciplinary and cosmopolitan reflection on several key considerations: what questions should be asked, whose views must be heard, what imbalances of power should be made visible, and what diversity of views exist globally.

We agreed that more crosstalk is needed between people representing different disciplines, political cultures and normative frameworks — so that approaches currently taken for granted can be tested and recalibrated in the light of alternative and dissenting perspectives. A new global forum, grounded in a commitment to hospitality and friendship towards unfamiliar, possibly upsetting ways of thought, would encourage people to build a rapport and so begin to engage more meaningfully with one another.

"Free enquiry, the lifeblood of science, does not mean untrammelled freedom to do anything."

To these ends, the global observatory we imagine would fulfil three functions.

First, it would serve as a clearing house. It would consolidate and make universally accessible the global range of ethical and policy responses to genome editing and related technologies. These responses would include relevant literature, and position statements from civil-society groups, especially from the global south. The network would also report on activities and outputs of formal bioethics bodies, such as the Nuffield Council on Bioethics in the United Kingdom or the German Ethics Council, professional societies such as the American Society for Reproductive Medicine, and intergovernmental agencies, such as the Council of Europe and the World Health Organization.

Second, the observatory would enable the tracking and analysis of significant conceptual developments, tensions and emerging areas of consensus around gene editing. It would broaden the focus beyond the technical pros and cons of gene editing to a richer range of questions and concerns that tend to be overlooked.

Studies of the social dynamics of international collaborations — from setting research agendas to the allocation of intellectual-property rights — could help to reveal the hidden power imbalances in science that are likely to influence who benefits from gene-editing research, as well as who does not. Likewise, the material gathered in the global observatory would give us a more detailed view of the biological futures people actually want for themselves and their societies. For instance, it could shed light on differing perceptions of social and biological relationships, such as ideas of disability and disease, across cultures.

Third, the observatory would serve as a vehicle for convening periodic meetings, and seeding international discussion informed by insights drawn from data collection and analysis.

To be effective in all three dimensions identified, those involved must reject the rhetoric of a competitive race in international science. The fixation on 'winning' should be replaced with deeper reflection on the purposes of technological change⁷. Analysis of the contexts in which the narrative of winners and losers emerges should itself be part of the work of the observatory, as should its effects on the course of scientific research.

REFRAMING THE QUESTIONS

If successful, the observatory we propose would alter the way problems are framed and expand the idea of a "broad societal consensus".

In current bioethical debates, there is a tendency to fall back on the framings that those at the frontiers of research find most straightforward and digestible. This move



The International Summit on Human Gene Editing was held in Washington DC in 2015.

comes at great cost. If the ethical stakes of human germline genome editing are limited to questions of physical safety, for example, then the technical evaluation of particular biological endpoints (for instance, off-target effects) might offer sufficient answers⁸. But such a focus short-circuits the central question of how to care for and value human life, individually, societally and in relation to other forms of life on Earth.

Likewise, the goals of consensus must go beyond merely agreeing on whether particular applications of genome editing are acceptable or unacceptable. Deliberation is insufficient if the conversation is too quickly boxed into judgements of the pros and cons, risks and benefits, the permissibility or impermissibility of germline genome editing, and so on.

Such an approach neglects important background questions — who sits at the table, what questions and concerns are sidelined, and what power asymmetries are shaping the terms of debate. When it comes to shaping the future of humanity, those neglected issues are just as important as the concerns of people poised to radically remake it. Indeed, consensus might even mean agreeing not to proceed with some research until a more equitable approach to setting the terms of debate is achieved⁹.

Cosmopolitanism, in styles of thought and in cultural intelligence, is not merely an aspiration for the proposed observatory; it should be integral to the network's way of working. Success will ultimately depend on whether those leading the initiative have the skill and sensitivity to manage cross-disciplinary and cross-cultural conversations, and are backed by the knowledge and

networks needed to sustain an infrastructure that facilitates these conversations.

LOOKING AHEAD

The observatory would not seek to engage in a race against science. Its purpose is more to engender robust, sustained conversation about the limits and directions of research. The pace of current research might well bring about some near-term interventions that humanity has not consented to, such as the creation of an edited child. Far from rendering international deliberation moot, such a step would only underscore the need for meaningful cosmopolitan thinking.

Our hope is that the observatory would begin to shift entrenched habits of thought beyond those directly influencing gene-editing research. Indeed, because the issues that the observatory would illuminate reach far beyond narrow questions about particular technologies and associated risks and benefits, its work should enrich and deepen debate around biotechnology more broadly.

All too often, scientists and others have tended to circumscribe debate about human genetic engineering on the premise that, until the technical capability does exist, it is not necessary to address difficult questions about whether such interventions in human life are desirable¹⁰. For example, even as scientists are applying gene editing to human embryos in the lab, the argument that the technology is too risky for clinical use serves as an excuse to delay the hard work of thinking through the technology's wider ramifications.

These tendencies to delimit and delay debate leave exploratory research largely unquestioned. The effect is that scientific developments, once they are realised, seem

to have been inevitable and outside our control, even though they are the products of scientists' choices. Questions of value then seem largely reactive, even futile.

Thus, a big challenge will be to ensure that entry cards to the observatory are not dictated by dominant cultural views about what constitutes relevant moral or technical competence. Profound and long-standing traditions of moral reflection risk being excluded when they do not conform to Western ideas of academic bioethics. But as the Intergovernmental Panel on Climate Change discovered through its climate assessment exercises, formal scientific training cannot be the only criterion by which to decide whose voices should be heard in an inclusive global forum. Equally, care must be taken to ensure that participation is not preferentially given to those who are the most vocal or most polarized on the issues.

Free enquiry, the lifeblood of science, does not mean untrammelled freedom to do anything. Society's unwritten contract with science guarantees scientific autonomy in exchange for a research enterprise that is in the service of, and calibrated to, society's diverse conceptions of the good. As the dark histories of eugenics and abusive research on human subjects remind us, it is at our peril that we leave the human future to be adjudicated in biotechnology's own "ecclesiastical courts".

It is time to invite in voices and concerns that are currently inaudible to those in centres of biological innovation, and to draw on the full richness of humanity's moral imagination. An international, interdisciplinary observatory would be an important step in this direction. ■

Sheila Jasanoff is professor of science and technology studies at the Harvard Kennedy School, Cambridge, Massachusetts, USA.

J. Benjamin Hurlbut is associate professor of biology and society at Arizona State University, Tempe, Arizona, USA.
e-mails: sheila_jasanoff@harvard.edu; bhurlbut@asu.edu

1. Baltimore, D. *et al.* *Science* **348**, 36–38 (2015).
2. National Academies of Sciences, Engineering, and Medicine. *On Human Gene Editing: International Summit Statement* (Organizing Committee for the International Summit on Human Gene Editing, 2015).
3. Jasanoff, S., Hurlbut, J. B. & Saha, K. *Issues Sci. Technol.* Fall, 25–32 (2015).
4. Sneed, O. C. *UC Davis Law Rev.* **43**, 1529–1604 (2009).
5. Steinfels, P. *Hastings Center Report* **6**(3), 21–25 (1976).
6. Culliton, B. J. *Science* **192**, 451–453 (1976).
7. Baylis, F. *Nature Hum. Behav.* **1**, 0103 (2017).
8. National Academies of Sciences, Engineering and Medicine. *Human Genome Editing: Science, Ethics, and Governance* (National Academies Press, 2017).
9. Jasanoff, S. *The Ethics of Invention: Technology and the Human Future* (Norton, 2016).
10. Hurlbut, J. B. *Experiments in Democracy: Human Embryo Research and the Politics of Bioethics* (Columbia Univ. Press, 2017).



ILLUSTRATION BY MARINA MUIN

Don't wait for an outcry about gene editing

The breadth of social and moral questions raised requires a new architecture for democratic debate, insists **Simon Burall**.

Over the past three years, thousands of articles have been published about editing genes and genomes. Apart from a public dialogue run by the Royal Society at the end of last year, there's been little attempt to engage the public on the implications of the technology in a way that could alter the decisions of scientists and policymakers. Indeed, concern about the lack of effective public engagement has motivated several workshops, including one by the inter-governmental Organisation for Economic Co-operation and Development (OECD).

If the history of public engagement surrounding other recent scientific innovations is a guide, efforts to explain the science behind gene editing will intensify, such as through news stories, at science festivals, in public lectures and in museums. And there will be a rash of small, disconnected workshops

involving members of the public that are designed to inform specific policy decisions.

If this is all that happens, scientists and policymakers will be ill prepared for the public debate that will almost certainly erupt as applications proliferate.

Instead of waiting for an outcry about a specific potential application, governments and key stakeholders must design a new kind of engagement.

GROWING AWARENESS

Since the 1980s, and especially over the past 10–15 years, there has been a steady rise in efforts to engage people in areas of complex and potentially controversial science and technology.

Approaches vary, but they generally involve specialists relaying information to small groups of the public. In Denmark in

1989, for instance, 15 members of the public, of various ages, social classes and ethnicities, talked to 15 experts about gene sequencing¹.

And in 2011, Sciencewise, a UK government programme that I directed, sponsored a public dialogue on experiments involving the transfer of DNA between animal and human cells, or the culturing of the two cell types together². Workshops involving scientists, policymakers and 70 members of the public, drawn from all walks of life, were accompanied by a survey of more than 1,000 people.

These kinds of processes can link different perspectives to policy. They're especially effective for advances that are likely to affect a limited number of people or domains, for instance, because of costs or special interests (as in rare diseases).

In the Danish case, participants provided a thoughtful response. They welcomed

screening for families affected by genetic diseases, but they expressed caution about the widespread use of genomics for diseases for which no cures exist, and the possibility of employers or health-insurance companies screening potential workers or applicants. And according to the organizers of the event, these concerns led to the Danish Parliament in 1997 prohibiting genetic screening for health-insurance purposes in 1997.

In the United Kingdom, too, most participants in the Sciencewise dialogue were in favour of research that could help to improve human health, as long as that research was robustly regulated and animal welfare was taken into account. Again, the engagement process influenced policy. It fed into the Guidance on the use of Human Materials in Animals, published in 2016 by the UK Home Office, which must be followed for experiments involving human and animal material to be licensed.

BROAD SCOPE

In my view, gene editing is likely to provoke a much broader set of social and moral questions, and therefore requires a different approach. The techniques are relatively cheap and easy to deploy. Like artificial intelligence, gene editing could radically alter almost every domain of life, including human health, plant and animal farming practices and the industrial production of drugs and materials.

A first step towards building an engagement structure for something of this scope is thinking about the public in a different way.

When pressed, most scientists and policymakers understand that 'the public' isn't an undifferentiated mass. Yet most engagement processes are designed on the premise that there is a set of unchanging perspectives out there waiting to be identified. They happen in spaces that the public is invited into, they are generally one-off or short and they involve little or no follow-up.

In reality, some people will have very limited knowledge; others will have a more sophisticated understanding. Those with certain values, politics and beliefs might be enthusiastic at the outset; others antagonistic. Still others, such as those belonging to patient groups or who are active in local conservation projects, might be interested but unaware of the technology. Crucially, most will already be embedded in an array of inter-linked networks — for example, as members of farmers unions, parent-toddler groups or activist organizations such as Greenpeace.

These existing networks offer numerous channels for widening awareness about gene editing and for sparking more-nuanced conversations about its possible impacts. They also offer policymakers and scientists windows onto what people are already saying (or not) about the technology.

The kind of engagement model I propose

would require government support and impetus from 10–15 organizations initially, to include the multitude of actors invested in and concerned about gene editing — perhaps with a single coordinating body. (In the United Kingdom, an institution such as the Nuffield Council on Bioethics could, in principle, mobilize others.)

Participating organizations would include those with declared interests in gene editing, such as research groups, the World Health Organization, national farmers unions, activist groups, pharmaceutical and agricultural companies and institutions focused on development, such as the Gates Foundation. They should also include groups that sit between stakeholders with a direct interest and the public, such as civil-society groups concerned with disability rights or the environment.

Such a consortium could first commission social scientists and others to use network analyses to map out which communities seem to be unaware of, or at least unengaged in, the wider societal debate about gene editing; who is already discussing it; and who is likely to want to know about it.

Ultimately, the consortium would have two functions: connecting people to the science and policy debates, and connecting scientists and policymakers to other people.

"In reality, some people will have very limited knowledge; others will have a more sophisticated understanding."

To achieve the former function, mapped networks would need to be given evidence-based information about both the science and policy in various forms and at different levels of

detail. And misconceptions and falsehoods would need to be corrected.

This could involve the consortium engaging on social media, producing videos, communicating with the media or providing information to relevant groups. As part of this effort to disseminate information and promote debate, the consortium could also connect communities with shared concerns, or individuals who are likely to share concerns. This could be done informally, through social media, say, or formally through workshops.

Crucially, the consortium would need to see itself as an active node in an existing network — one that helps to facilitate cross-talk and promote debate more broadly.

Once a critical mass of people are engaging with the social, political, moral and economic implications of gene editing (which will happen whether or not governments develop an effective engagement architecture), the societal debate will move fast. Thus, the selected organizations would need to communicate the range of views to policymakers and scientists, and correct the misconceptions

that these groups frequently have about people's perspectives on the development of a technology — that they will inevitably reject it or never understand it.

In the longer term, this effort could guide the commissioning of more-conventional engagement processes around specific policy decisions. Such efforts are likely to be much more effective if built on ongoing network-based engagement.

TOUGH TACTICS

What I am proposing will be difficult.

Researchers and policymakers may be reluctant to commit to a new architecture for developing a debate that no one is able to control. Many are sceptical about the value of public involvement in making decisions about science and technology³. Certainly, specialists will need to demonstrate humility, a clear understanding of the bounds of their own expertise and an appreciation of the knowledge, experience and values that members of the public can bring.

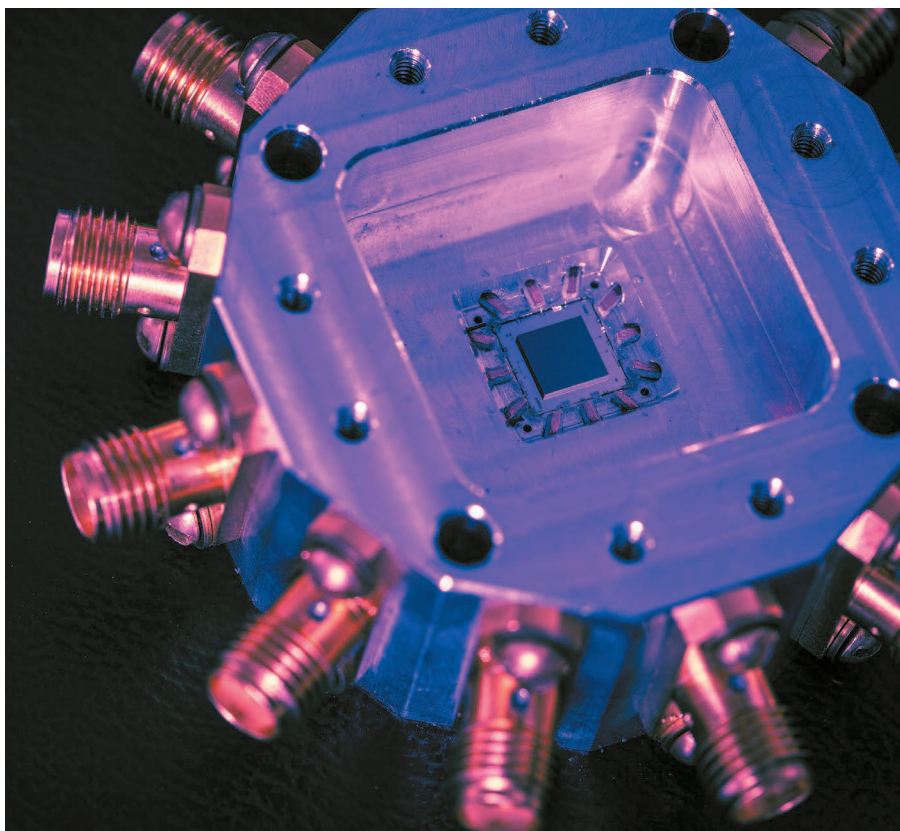
The most viable way forward might be for governments to identify key stakeholders who share concerns about the public debate becoming dominated by people with competing views talking past each other. Together, these organizations could produce a statement of intent and start to build political support for the new entity, its role and membership.

This is not uncharted territory. Social scientists at the Science, Society and Sustainability (3S) Group at the University of East Anglia, UK, mapped public engagement with issues on energy use between 2010 and 2015. Then, last October, the team proposed an architecture (similar to the one I set out here) for the energy and climate-change debate⁴.

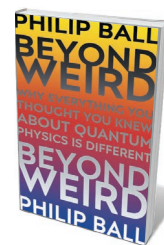
Finally, the model I propose will require significant investment of time and money (in the region of US\$700,000 to \$1.5 million per year). Yet, as many experiences with genetically modified crops have illustrated, simply trying to convince people that the science is safe, and that they should accept the applications that emerge, can be much more costly. ■

Simon Burall is a senior associate at the UK charity *Involve* in London, which provides advice to the Sciencewise programme.
e-mail: simon@involve.org.uk

1. Danish Board of Technology, Consensus Conference on the Application of Knowledge gained from Mapping the Human Genome (Danish Board of Technology, 1989); available at go.nature.com/2oxtrj1
2. The Academy of Medical Sciences. Animals Containing Human Material (The Academy of Medical Sciences, 2011).
3. Grove, J. 'Populism and polarisation "threaten science", Nobel laureates say' *Times Higher Education* (31 August 2017).
4. Chilvers, J., Pallet, H. & Hargreaves, T. Public Engagement with Energy: Broadening Evidence, Policy and Practice (UK Energy Research Centre, 2017).



A computer chip created by Google Research to manipulate data using quantum bits (qubits).



Beyond Weird:
Why Everything
You Thought
You Knew About
Quantum Physics
Is Different

PHILIP BALL

Bodley Head: 2018.

QUANTUM PHYSICS

A trek through the probable universe

Natalie Wolchover enjoys Philip Ball's grapple with slippery questions about the quantum revolution.

It takes care to handle the paradox of describing something that no one truly understands. The fastidious, plainspoken science writer and former *Nature* editor Philip Ball mostly achieves this ambition in his 23rd book, *Beyond Weird*, a clear and deeply researched account of what's known about the quantum laws of nature, and how to think about what they might really mean.

Quantum theory, developed in the early twentieth century by Max Planck, Albert Einstein, Niels Bohr and other luminaries, strips 'reality' of its usual meaning by seeming to say that nothing is certain about atoms or elementary particles until they are measured. A particle's location, for example, is a matter of probability. It is defined by a spread-out, abstract 'wavefunction' until the act of looking for the

particle somehow collapses the wavefunction to one spot, inducing a single reality. We inhabit, as theoretical physicist John Wheeler said, a "participatory universe" — and confusion stems from there. Ball tells us that quantum theory is "cognitively impenetrable". Yet he convincingly argues that the questions about it have sharpened and even changed profoundly over time — progress, in lieu of answers.

Along with the historic discoveries, Ball brings readers up to speed on today's "quantum renaissance". This active intellectual period is fuelled by quantum-computing research and the rise of quantum information theory, pioneered by researchers including David Deutsch, Peter Shor and Charles Bennett. Quantum mechanics is now seen as, more than anything, a set

of rules about how information can be shared and processed. That, Ball says, is why quantum computing has proved so stimulating: what's possible and impossible to compute "follow from the same rules that govern what is knowable and what is not". Additionally, physicists, starting with Dieter Zeh in the 1970s and Wojciech Zurek in the 1980s, have developed a decent

understanding of the quantum-classical divide: the reason particles can exist in superpositions of many possible states, but cats are only ever dead or alive. Classical physics is now seen as emerging from underlying quantum laws because of 'decoherence'. That is, brushes with the environment cause quantum superpositions to lose coherence as information about them leaks out — and the bigger the system, the faster it happens.

Modern theoretical studies and exquisitely controlled experiments have revealed the newly important role of principles such as information causality, which puts a hard limit on the amount of information that can be deduced about a quantum system after it has been measured. Researchers in quantum reconstruction are taking the new insights and going back to basics, looking for the minimal set of axioms from which the full range of quantum behaviours emerges.

Ball is an intelligent guide. The weirdness is dizzying at times, however, and landmark discoveries might be missed on his meandering tour of non-intuitive phenomena and their implications. He gives a rather low-impact description of the double-slit experiment. Yet this is arguably the most instructive demonstration of quantum weirdness: a particle behaves like a wave that passes through both slits in a screen and interferes with itself on the other side, producing peaks of probability in the places where it might then be found. Also lacking oomph is the description of Bell's theorem, the 1964 watershed that enabled physicists to prove experimentally that entangled quantum particles exhibit 'non-locality'. That finding upended the classical belief that things must exist in places.

Other passages are riveting, such as Ball's elaboration of the fascinating debate surrounding the meaning of the wavefunction. As he writes, efforts to crack quantum mechanics have forced scientists to examine ancient philosophical questions about the nature of reality, knowledge and existence. Do quantum wavefunctions exist, for instance, or do these probabilistic descriptions of

the world reflect our own ignorance?

Ball cites the Copenhagen interpretation, the hard line taken by deep-thinking Danish physicist Niels Bohr, who said, “There is no quantum world. There is only an abstract quantum physical description.” Bohr also noted that physics concerns not how nature is, but “what we can say about nature”. (Theoretical computer scientist Scott Aaronson has defined Copenhagenism as “shut up and calculate except without ever shutting up about it”.) Ball sees Bohr’s perspective as frustratingly elusive.

The competitors hardly fare better. The de Broglie–Bohm interpretation, which imagines particles guided by invisible pilot waves, is, in Einstein’s words, “too cheap”. Stochastic-collapse models equip wavefunctions with extra components that randomly choose their states at the moment of measurement — “a bodge”, according to Ball. The consistent-histories interpretation, which says that anything that happens has to be consistent

“Ball is at his liveliest when incisively critiquing the many-worlds interpretation.”

ent with what came before, he sees as useful, although he also says that it fails to supply “a physical picture”.

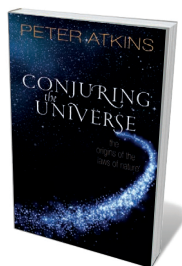
A theory known as QBism takes an even harder line than Copenhagenism, interpreting quantum wavefunctions as personal beliefs — but it does claim compellingly that reality, rather than being rigid and deterministic, exhibits creativity or novelty, “almost a lawlessness out of which laws can arrive”.

Ball is at his liveliest when incisively critiquing the many-worlds interpretation. This popular hypothesis holds that every measurement sends the Universe along branching paths, proliferating the number of parallel realities. Logical problems arise, he argues, when you assume that every quantum probability plays out — including probability itself becoming meaningless.

No angle on the quantum world feels satisfactory, but many add insight. There’s a sense of growing perspective in a haze of confusion. *Beyond Weird* distinguishes itself by giving life to the ideas, rather than just to Einstein, Bohr, Erwin Schrödinger and the other colourful quantum pioneers. Their discoveries and quandaries about spooky action at a distance and feline superpositions were so deliciously disorienting that they’ve been endlessly reshaped by other authors. Ball brings them all into the prose, but in service of the science. “Quantum theory had the strangest genesis,” he says. “Its pioneers made it up as they went along. What else could they do?” ■

Natalie Wolchover is a physics reporter and senior writer for Quanta Magazine. She lives in Brooklyn, New York.
e-mail: nwolchover@quantamagazine.org

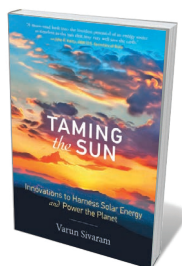
Books in brief



Conjuring the Universe: The Origins of the Laws of Nature

Peter Atkins OXFORD UNIVERSITY PRESS (2018)

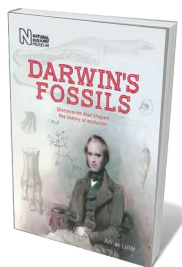
It’s rare to find a study of physical laws that is also a bravura display of rarefied humour and experiential depth; but such is this gem by chemist Peter Atkins. Looking at both ‘inlaws’ and ‘outlaws’ — the deep structural laws of the cosmos and minor elaborations such as Robert Hooke’s law of elasticity — Atkins explores the conservation of energy; laws related to temperature, electricity and magnetism; and fundamental constants such as the speed of light. He also shows, in one mind-bending chapter, how something really did come from nothing at the moment of cosmogenesis.



Taming the Sun

Varun Sivaram MIT PRESS (2018)

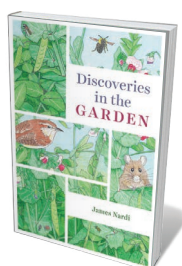
In 2017, solar power became the fastest-growing new-energy source. Yet as energy-policy specialist Varun Sivaram details in his global survey, that surge will be unsustainable unless countries invest in solar innovation on three fronts: financial (to amass investment); technological (to capture and store solar energy); and systemic (to re-engineer electricity grids). Solar power will also have to grow 30-fold by 2050. Sivaram includes a raft of case studies, from current research on the photovoltaic materials called perovskites to Off Grid Electric, a start-up aiming to electrify swathes of Africa by 2019.



Darwin’s Fossils

Adrian Lister NATURAL HISTORY MUSEUM (2018)

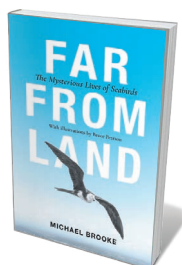
A “young man of promising ability, extremely fond of geology”. That accolade given to Charles Darwin by the captain of the HMS *Beagle*, Robert FitzRoy, propelled the fledgling scientist on the ride of his life. As naturalist on the ship’s 1831–36 voyage to South America, Darwin came into his own as a geologist, illuminating the formation of coral reefs and continental uplift. And, as palaeobiologist Adrian Lister details in this deft, beautifully illustrated account, no less important were Darwin’s discoveries of gargantuan mammal fossils — including the 1.5-tonne giant sloth *Myodon darwini*.



Discoveries in the Garden

James Nardi UNIVERSITY OF CHICAGO PRESS (2018)

Biologist James Nardi urges us to recall that it’s a jungle out there — in the back garden, that is. His companion guide to garden science is a learned romp through plant biology; solar energy and soil nutrients; the movement of vines; and “fellow gardeners”, from single-celled protozoans to beetles. Here, too, are experiments on the basics, such as photosynthesis, and a wealth of stunning images. These range from a prickly study of weed burrs to a scanning-electron-microscope shot of trichomes, hairs that stud tomato leaves like “dark lollipops” and give the plant its pungent odour.



Far from Land: The Mysterious Lives of Seabirds

Michael Brooke PRINCETON UNIVERSITY PRESS (2018)

For 40 years, ornithologist Michael Brooke has traversed the seas in pursuit of seabirds — from Alexander Selkirk Island west of Chile to the Shiant Isles of the UK Hebrides. His tour of the ecological highlights is gripping in scope and in granular detail gleaned from observations of puffins, penguins, fulmars and more. Brooke begins at fledging and glides on through migration, navigation, breeding and myriad threats facing these avian wonders, from introduced predators and light pollution to overfishing and ocean plastic. **Barbara Kiser**



Real and fictional specimens appear side by side in a project called *Die Sammlung* (The Collection).

ART

Fake views

Anthony King enjoys an exhibition of hoaxes, frauds, simulacra and substitutes.

As fake news continues to be big news, the Science Gallery Dublin nudges visitors to look anew at the value of distortion with its latest show, FAKE. The exhibition “probes our society’s flexibility with fact and fiction”, notes head of programming Ian Brunswick. “We want visitors to consider contexts where fake is not such a bad thing.” After all, phony things that succeed become a substitute: think faux fur or synthetic insulin.

Fakery of data in research is ruinous, but art has a licence to playfully subvert or embellish the real. As literary luminary Oscar Wilde, born around the corner from the gallery, had it: “Lying, the telling of beautiful untrue things, is the proper aim of art.”

In *The Art of Deception*, designer Isaac Monté and evolutionary biologist Toby Kiers present ghost hearts: six animal hearts stripped of their cells to leave behind sterile protein scaffolds. Repopulating such frameworks with human cells could one day help to save lives, enabling a suboptimal organ to be replaced. They offer a blank slate for ingenious engineering. “This is both its power and its danger,” says Kiers. One of the hearts has been dipped in hot wax, one recast with foam, one dried like jerky in an airflow cabinet. That enhanced alienness reminds us that we must

FAKE
Science Gallery
Dublin.
Until 3 June 2018.

deceive our bodies into accepting organs from other species. Meanwhile, artist Heather Beardsley playfully contrasts display conventions in museums in *Die Sammlung* (The Collection), a pleasing menagerie of real and fictional specimens in antique jars. I picked out a salamander, coral and spider crab, but with these were weird creatures of dubious phylogeny. Close inspection is needed to separate the zoological shams from specimens on loan from the National University of Ireland Galway.

Several exhibits centre on food. Visitors can sample titbits in *FAUX Foodmongers*, from krab sticks to vegan cheese. Our notions of authenticity and naturalness are probed by artist Crystal Bennes in $\neq C_8H_8O_3$, a chance to sample four vanilla-flavoured tablets: natural vanilla and three vanillins, synthesized from a petrochemical product, a wood by-product and fermented yeast. Only 1% of vanilla flavour in foods is derived from vanilla pods, yet supply still can’t satisfy demand. Recycled spruce wood may be sustainable, but the pod wins on taste.

Patricia Pisanelli’s *Stretching Cheese*, hanging downstairs, is a collage of slices of

the stuff, reminding us that products with a cheese content of just 51% can be sold as cheese. Devon Ward and Oron Catt’s *Vapour Meats* is a plastic see-through head mask through which meat flavours can be inhaled. A motion sensor triggers a nozzle that sprays a chemical vapour into the mask. I wanted to test its meaty bona fides, but university vaping regulations quashed the idea: I’m left wondering whether the odours are fake, or deliver the fatty umami-drenched whiff of the real thing.

Other works explore the deception rife in the animal world. Cuttlefish, for example, rapidly change colour and texture in response to their environment — camouflage that underlies *Cuttle 61*. Artist Ryuta Nakajima decorated aquaria with reproductions of iconic visual art. His photos show how the creatures responded — for instance, turning uniform grey with eye spots to blend with Pablo Picasso’s *Ma Jolie*. Ecologists can deploy innovative deception for their own ends. In a glass box sits *Faux Frog*, artist Barrett Klein’s collection of artificial amphibians. These were used in field studies of signalling and mate selection by female *Túngara* frogs in Panama.

We, too, can be gulled — in some realms more readily than in others. People can spot fake food, flowers or fruit within 40 milliseconds, notes curator and psychologist Fiona Newell. But our antennae for faked photos might be less acute. A sample from a fake-news exhibition launched last year at the UK National Science and Media Museum in Bradford dates back as far as 1911. I was particularly struck by a newspaper image of two child victims of a crime, set against a fake seaside background. (The paper modified the image — perhaps to ramp up the emotional impact by showing the children in a moment of childhood bliss.)

The ease of hoaxing on social media is clear from *Synthesizing Obama*. A realistic video shows former US president Barack Obama speaking, but lip-synced to a speech he never made. Realistic faked video and audio develops apace, notes Brunswick; machine-learning techniques have also been harnessed to make short fake videos.

FAKE’s aim is not to frighten people into thinking they are drowning in duplicity, but to probe why we value authenticity. Says Brunswick: “We can’t let the fake-news scandal dominate our entire cultural, scientific and artistic discussion.” Instead, this show explores how questionable truthfulness, deception and ingenious subterfuge tease us. “If people come away with more questions than answers, about fact and fiction and the fuzzy in between,” concludes Brunswick, “we’ve done our job.” ■

Anthony King is a writer based in Dublin. He tweets @AntonyJKing. e-mail: anthonyjking@gmail.com

CORRECTION

The exhibition review 'Fake Views' (*Nature* **555**, 442; 2018) erroneously described the artist Ryuta Nakajima as a woman.

Correspondence

Bitcoin mining is not uneconomic

In my view, Spyros Foteinis's arguments against bitcoin's burgeoning technology on the grounds of its carbon footprint are economically unsound (*Nature* 554, 169; 2018). He compounds the common misperception that cryptocurrency mining is inherently wasteful.

Foteinis estimates that the combined annual electricity consumption due to bitcoin and ethereum mining is 80% that of Greece. But Bitcoin and ethereum last year together generated wealth of US\$275 billion (see go.nature.com/2drhvf). Given that the 11 million inhabitants of Greece generated wealth of almost US\$205 billion in the same year (see go.nature.com/2ha6mqg) and would have consumed much more than just electrical energy in the process, the mining of cryptocurrency seems considerably less wasteful.

The only reform required here is for all nations to move away from fossil-fuel-based electricity generation to nuclear or other sustainable alternatives. The specific association of cryptocurrencies with this unconditional imperative distracts from the real problem.

Stuart Wimbush Victoria University of Wellington, New Zealand.
stuart.wimbush@vuw.ac.nz

Inspire lead authors from global south

Dyna Rochmyaningsih argues that the position of authors from the 'global south' on papers with Western scientists could be unfairly affected by their limited access to funding (*Nature* 553, 251; 2018), citing our paper on a new species of orangutan (*Pongo tapaniensis*) in Indonesia as an example. In that case, however, an Indonesian (A.N.) is one of 4 lead authors, and listing of the 37

co-authors was decided on the basis of contribution rather than funding. Nevertheless, such collaborative standards are rare.

Authorship is too frequently a bargaining point to enable international scientists to work in developing countries or it is politicized, as seems to be happening in Indonesia. National and international scientists from the global south are still less likely to have their papers published (K. A. Wilson *et al.* *PLoS Biol.* 14, e1002413; 2016) and cited (E. Meijaard *et al.* *Conserv. Biol.* 29, 920–925; 2015). These biases need to be addressed.

Access to funding is not the key constraint for aspiring lead authors from the global south (see also *Nature* 554, 415–416; 2018); scientific capacity is. They need more and better instruction on how to lead the conceptualization, implementation, analysis and write-up of the research. As well as participation in international research, this requires training in scientific leadership, publishing culture and in English reading and writing skills. Once they become science leaders, this will rapidly translate into lead authorship.

Anton Nurcahyo Australian National University, Canberra, Australia.

Erik Meijaard University of Queensland, Brisbane, Australia.
anton.nurcahyo@anu.edu.au

Food safety: risk of glyphosate flagged

As journalists, we have dug into the controversy over the suspected carcinogenicity of the herbicide glyphosate (see B. Url *Nature* 553, 381; 2018). We searched hundreds of internal documents from its manufacturer, Monsanto, released for pending lawsuits (see go.nature.com/2tfpbwy). We feel that our and others' findings (N. Ghisi *et al.* *Chemosphere* 145, 42–54; 2016) justify concern over

the expertise used by the regulatory agencies to evaluate the safety of glyphosate.

From these documents, we formed the impression that Monsanto's in-house toxicologists were concerned about the upcoming evaluation of their flagship product by the International Agency for Research on Cancer, and that they had anticipated the result before the agency classified the product as 'probably carcinogenic to humans' in March 2015. In 1999, the company had commissioned a confidential external expert review of the evidence, which suggested that glyphosate was mutagenic.

In 2016, the European Food Safety Authority declared the evidence on glyphosate's carcinogenicity for humans to be "very limited" (see go.nature.com/2tjuql1), partly on the basis of a large prospective study, the Agricultural Health Study (AHS; see also G. Andreotti *et al.* *J. Natl Cancer Inst.* <http://dx.doi.org/10.1093/jnci/djx233>; 2017). Yet Monsanto's chief epidemiologist wrote in an internal report in 1997 that the AHS protocol rendered estimates of exposure to pesticides inaccurate, potentially masking associations between exposure and disease.

Stéphane Foucart, **Stéphane Horel** Le Monde, Paris, France.
foucart@lemonde.fr

Food safety: change testing regime

I agree with the director of the European Food Safety Agency (EFSA) that the historical mistake of mandating industry to test their own products should not be repeated (B. Url *Nature* 553, 381; 2018). The agency's own record badly shook public trust in 2010, for example, when it overrode

800 or so academic studies on the chemical bisphenol A and its risk to health and the environment — relying instead on the results of four industry-sponsored protocol studies (see go.nature.com/2fi1vcs).

Government-funded academic studies have no vested interest in manipulating data. The best take into account factors such as the cumulative risk of exposure (for example, to dozens of pesticide residues at the same time), do not assume safe levels of carcinogens, test experimentally rather than dismiss low-dose effects of chemicals, and incorporate epidemiology into their evaluations.

On the weedkiller glyphosate, the 'Monsanto papers' (see go.nature.com/2tfpbwy) indicate that the main problem has been the manipulation of data in industry-sponsored studies. In my view, the only role for industry is to contribute to the cost of safety-evaluation studies.

A full reform of EFSA panels would help to restore public trust in the agency's work. For example, a stricter conflict-of-interest policy would promote independence and objectivity, and including more academic scientists would boost scientific insight and raise scientific standards.

Hans Muilerman Pesticide Action Network, Brussels, Belgium.

hans@pan-europe.info

*Competing financial interests declared, see go.nature.com/2poqyxv

CONTRIBUTIONS

Correspondence may be sent to correspondence@nature.com after consulting the author guidelines at <http://go.nature.com/cmchno>.

Stephen Hawking

(1942–2018)

World-renowned physicist who defied the odds.

When Stephen Hawking was diagnosed with motor-neuron disease at the age of 21, it wasn't clear that he would finish his PhD. Against all expectations, he lived on for 55 years, becoming one of the world's most celebrated scientists.

Hawking, who died on 14 March 2018, was born in Oxford, UK, in 1942 to a medical-researcher father and a philosophy-graduate mother. After attending St Albans School near London, he earned a first-class degree in physics from the University of Oxford. He began his research career in 1962, enrolling as a graduate student in a group at the University of Cambridge led by one of the fathers of modern cosmology, Dennis Sciama.

The general theory of relativity was at that time undergoing a renaissance, initiated in part by Roger Penrose at Birkbeck College, London, who had introduced new mathematical techniques. These showed that generic gravitational collapse would lead to singularities — infinities that signal the need for new physics.

The implications for black holes and the Big Bang were developed by Hawking in a series of papers collated in the 1973 monograph *The Large Scale Structure of Space-Time* (Cambridge University Press), co-authored with George Ellis, a near-contemporary who had also been a student of Sciama. Especially important was the realization that the area of black holes' horizons ('one-way membranes' that shroud the singularities, and from within which nothing can escape) could never decrease. The analogy with entropy — a measure of disorder that likewise can never decrease — was developed further by physicist Jacob Bekenstein.

These findings gained Hawking election to the Royal Society in London in 1974, at the age of 32. By then, he was so frail that both movement and speech were difficult, and most of us suspected that his days in front-line research were numbered. But in that same year, he came up with his most distinctive contribution to science: Hawking radiation.

By linking quantum theory and gravity, Hawking showed that a black hole would not be completely black, but would radiate with a well-defined temperature that depended inversely on its mass (S. W. Hawking *Nature* **248**, 30–31; 1974). Black-hole entropy was more than just an analogy. The implication



was that the radiation would cause black holes to 'evaporate'. This process would be unobservably slow, except in 'mini-holes' the size of atoms — and these are thought not to exist. Yet Hawking radiation — and the related issue of whether information that falls into a black hole is lost or is somehow recoverable from the radiation — was a profound issue, and one that still engenders controversy among theoretical physicists. Indeed, theorist Andrew Strominger at Harvard University in Cambridge, Massachusetts, said in 2016 that one of Hawking's papers on the subject (S. W. Hawking *Phys. Rev. D* **14**, 2460–2473; 1976) had caused "more sleepless nights among theoretical physicists than any paper in history".

By the end of the 1970s, Hawking had been appointed to the Lucasian Chair of Mathematics at Cambridge (former incumbents include Isaac Newton and Paul Dirac); he held the post until he retired in 2009. During these years, in which his focus shifted to the quantum aspects of the Big Bang, the issue of information loss in black holes continued to challenge him.

In 1985, Stephen underwent a tracheotomy, which removed his already limited powers of speech. He was able to control a cursor on a screen and type out sentences — albeit with increasingly painful slowness (first with his hand, and eventually only with a cheek muscle). A

speech synthesizer processed his words and generated the androidal accent that became his trademark. In this way, he completed his best-selling book *A Brief History of Time* (Bantam, 1988), which propelled him to celebrity status.

Had Hawking achieved equal distinction in any other branch of science besides cosmology, it probably would not have had the same resonance with a worldwide public. As I put it in *The Telegraph* newspaper in 2007, "the concept of an imprisoned mind roaming the cosmos" grabbed people's imagination.

In 1965, Stephen married Jane Wilde. After 25 years of marriage, and three children, the strain of Stephen's illness and of sharing their home with a team of nurses became too much and they separated, divorcing in 1995. Jane wrote a book about their life together, *Travelling to Infinity* (Alma, 2008), and both she and Stephen were happy with the telling of their story in the 2014 film *The Theory of Everything* (although it elides and conflates Stephen's science). After a second, briefer marriage, Stephen was supported by an entourage of assistants, as well as his family.

Stephen remained remarkably positive throughout his life, despite the immense frustration that his condition clearly caused. He enjoyed theatre and opera trips, and he seemed energized rather than exhausted by his travels to all parts of the world, as well as by his regular trips to the California Institute of Technology in Pasadena. He retained robust common sense and a sense of humour, expressed forceful opinions, supported political causes and was happy to engage with the media, despite its insistent attention. His comments gained outsized attention even on subjects in which he was not a specialist, such as philosophy and the dangers of artificial intelligence.

Stephen's expectations when he was diagnosed dropped to zero; he said that everything that had happened since had been a bonus. And what a bonus — for physics, for the millions enlightened by his books and for the even larger number inspired by his achievement against all the odds. ■

Martin Rees is Astronomer Royal of the United Kingdom. He was a student in Dennis Sciama's research group at the University of Cambridge at the same time as Stephen Hawking.
e-mail: mjr@ast.cam.ac.uk

IAN BERRY/MAGNUM

Machine learning classifies cancer

Brain tumours are often classified by visual assessment of tumour cells, yet such diagnoses can vary depending on the observer. Machine-learning methods to spot molecular patterns could improve cancer diagnosis. [SEE ARTICLE P.469](#)

DEREK WONG & STEPHEN YIP

Accurate diagnosis is essential for appropriate disease treatment. A core technique used to diagnose brain cancer today is the microscope-based analysis of tumour samples on glass slides, termed histology. However, this requires the appraisal of subtle cellular alterations, which in some cases may lead to different classifications for a given sample by different individuals. Nowadays, technological developments enable vast amounts of molecular data to be obtained and assessed for a tumour without the need for such subjective diagnostics. Machine-based-learning approaches are being developed to aid the diagnosis of clinical samples, and on page 469, Capper *et al.*¹ report such a method for classifying brain tumours on the basis of molecular patterns.

In 1926, a publication entitled *A Classification of the Tumors of the Glioma Group on a Histo-Genetic Basis with a Correlated Study of Prognosis*² by neurosurgeons Percival Bailey and Harvey Cushing provided early insight into the development, cellular characteristics and clinical consequences of glioma, a type of cancer of the central nervous system (CNS). The book's title was prophetic and ambitious, given that the microscope-based diagnostic approach they advocated was not common then. The authors' ideas were ahead of their time — for example, the word 'histo-genetic' in the book's title points to a link between cellular changes and genetics. Bailey and Cushing's obsessive attention to detail allowed them to identify gross and microscopic tumour features that correlated with clinical outcomes, and the book reported the classification of 14 types of tumour.

Today, many brain tumours are identified by analysis of both histological and molecular features^{3–5}. The identification^{6,7} of biologically relevant, tumour-type-defining and clinically informative genetic alterations in brain tumours prompted the World Health Organization (WHO) to update its diagnostic guidelines for certain brain tumours in 2016 to recommend an integrative diagnostic approach that combines both histology and molecular information^{8,9}. However, diagnoses that rely predominantly on histology remain

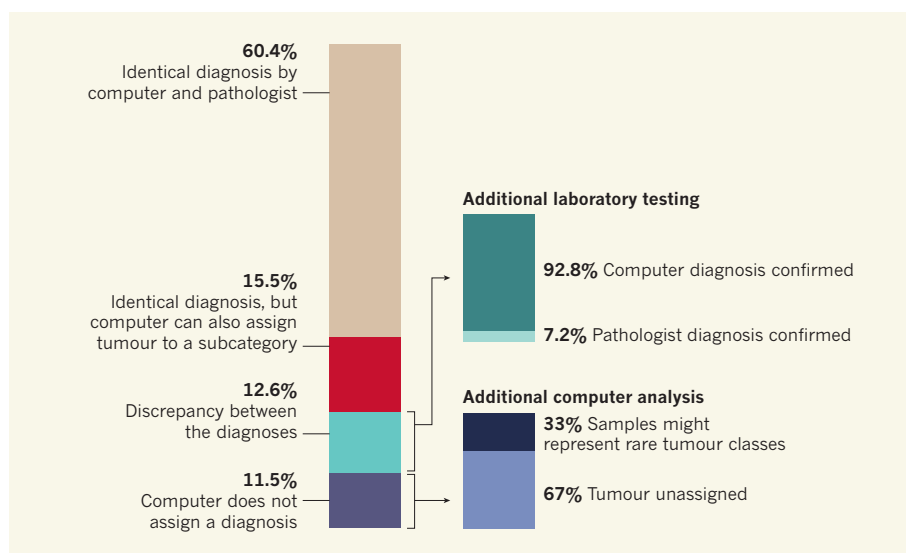


Figure 1 | Tumour classification using a machine-learning approach. Capper *et al.*¹ used a machine-learning approach to classify brain tumours on the basis of genome-wide patterns of a type of DNA alteration called methylation. The computer was trained using methylation data for tumour samples that had been diagnosed by pathologists using standard microscopy-based analysis or analysis of selected genes. After training, the computer was given 1,104 test cases. The authors compared the diagnoses made by the computer and by the pathologists. Although the machine was unable to diagnose all specimens, of the specimens that it classified, the machine-based diagnosis was more accurate or could assign tumours to more-specific subcategories than the classifications made by the pathologists.

common for many types of rare tumour, owing to a lack of molecular identifiers. Yet histological diagnoses face many challenges, including possible cellular variations in tumours that are a mosaic of cells containing different genetic alterations, or the fact that similar histological features can be shared by many different types of brain tumour. Questions remain about how well histological similarity reflects tumour similarity, given that tumours that have similar histology can progress in different ways, and tumours that have contrasting histology can progress in the same way.

A key development for histological analysis is the expansion of computational tools that allow machine-learning processes to analyse histological data^{10,11}. In this approach, a computer is 'trained' using a data set of sample images of tumours that have been classified by a physician. The computer uses the classification information to develop its own pattern-recognition criteria with which to identify tumour types. However, a challenge arises if

clearly defined diagnostic criteria for certain tumours are lacking, or if different types of tumour are histologically indistinguishable.

Capper and colleagues decided to focus on molecular information whose classification does not require complex visual assessments. They took a machine-learning approach to tumour classification based on changes in DNA methylation — the addition of methyl groups to DNA — and compared such diagnoses with those made by pathologists using histological analysis.

DNA methylation is a type of modification known as an epigenetic change. This category of alteration does not change the DNA sequence but can affect gene expression or cell fate. The role of aberrant DNA methylation and other epigenetic changes in cancer is becoming increasingly evident^{12,13}. In many cancers, the genome-wide pattern of epigenetic changes, known as the epigenome, can be substantially altered. For example, mutations in the genes *IDH1* or *IDH2* in gliomas

cause genome-wide dysregulation of DNA-methylation patterns that can be correlated with specific clinical outcomes¹².

Previous studies^{14–16} have highlighted the diagnostic advantages of profiling DNA methylation for certain types of brain tumour because — compared with histology or the testing of specific genetic alterations — an epigenome-wide analysis of DNA methylation offers an unbiased diagnostic approach. Yet routine epigenome-wide methylation profiling remains relatively uncommon for clinical diagnosis for several reasons, including: cost; sample requirements; a shortage of staff with the necessary data-analysis expertise; and the question of whether the findings would have implications for the clinical treatments used. However, some progress is being made. For example, techniques are now available to use DNA extracted from the most common type of chemically preserved tumour tissue on glass slides, called formalin-fixed, paraffin-embedded (FFPE) specimens.

The authors provided the computer with genome-wide methylation data for samples of almost every CNS tumour type classified by the WHO. The computer used supervised machine learning to recognize methylation patterns present in the pathologist-classified samples, as well as unsupervised machine learning, which involved the computer searching the data sets for patterns that it could use to assign samples into its own computer-generated classification categories.

After training, the computer could classify tumours into 82 distinct classes on the basis of specific methylation profiles. Only 29 of these corresponded to a specific tumour type as defined by the WHO and another 29 represented subclasses of the WHO-defined tumour types.

Yet perhaps the most interesting discoveries made by Capper and colleagues were tumour classifications that grouped together histologically similar types of tumour comprising more than one tumour type as classified by the WHO, or classifications of tumour types that did not match the WHO groupings. Such discoveries might provide insight into tumour similarities that are independent of tumour histology and could aid the development of treatment options or diagnostic tools.

The authors used the computer to classify 1,104 test cases of tumours that had been diagnosed by pathologists using standard histological or molecular techniques (Fig. 1). For 60.4% of these test cases, the computer-based classification was identical to the pathologist's classification, and for 15.5% of the test cases the computer and pathologist assigned the same type of tumour but the computer could also assign the tumour into a subclass. In 12.6% of the test cases, the computer diagnosis did not match the pathologist's diagnosis. Remarkably, further rigorous analysis of these cases — by, for example, gene sequencing — resulted in

the classification of 92.8% of these unmatched tumours being switched from the original clinical diagnosis to the computer-based classification. Moreover, 71% of the reclassified tumours were assigned to a different tumour grade, a recategorization that might have implications for prognosis or treatment. The remaining test cases (11.5%) could not be classified by the computer. Additional computational analysis indicates that one-third of the tumours in this group might represent rare tumours for which the computer had yet to encounter enough examples to generate a classification grouping.

Does Capper and colleagues' approach represent a probable future standard for tumour diagnosis, given the advantages, such as a low cost per sample that is comparable to that of standard cancer diagnostics; the compatibility with universally available FFPE material; and a website that facilitates data entry, analysis and tumour classification? And, if so, will histological analysis fall by the wayside?

Obtaining a comprehensive molecular profile of a tumour specimen is certainly useful, especially when combined with microscopic examination, and might be the way forward as medical treatments become ever-more personalized to the characteristics of an individual's tumour. However, for now, histology remains indispensable for disease classification because the standard approaches for specimen preservation and examination by microscopy offer the most accessible and universal entry point in the routine diagnostic workflow used in clinical laboratories worldwide. A disease can manifest itself in both molecular and cellular changes; therefore, an approach that integrates both molecular analysis and visual inspection might strengthen diagnostic capabilities.

Routine and widespread use of the platform developed by Capper *et al.* might not be practical for many laboratories at present, so the most likely immediate application of this technology would be in assessing cases with ambiguous histological characteristics. Nevertheless, Capper and colleagues' approach complements, extends and, in some cases, supersedes the tumour-diagnostic potential of microscopic examination. ■

Derek Wong and Stephen Yip are in the Department of Pathology and Laboratory Medicine, University of British Columbia, Vancouver, British Columbia V6T 2B5, Canada.

e-mail: stephen.yip@vch.ca

1. Capper, D. *et al.* *Nature* **555**, 469–474 (2018).
2. Bailey, P. & Cushing, H. *A Classification of the Tumors of the Glioma Group on a Histo-Genetic Basis with a Correlated Study of Prognosis* (Lippincott, 1926).
3. Cancer Genome Atlas Research Network. *N. Engl. J. Med.* **372**, 2481–2498 (2015).
4. Eckel-Passow, J. E. *et al.* *N. Engl. J. Med.* **372**, 2499–2508 (2015).
5. Sturm, D. *et al.* *Cell* **164**, 1060–1072 (2016).
6. Yan, H. *et al.* *N. Engl. J. Med.* **360**, 765–773 (2009).
7. Sturm, D. *et al.* *Cancer Cell* **22**, 425–437 (2012).
8. Louis, D. N., Ohgaki, H., Wiestler, O. D. & Cavenee, W. K. (eds) *WHO Classification of Tumours of the Central Nervous System* 4th edn (International Agency For Research on Cancer, 2016).
9. Aldape, K., Nejad, R., Louis, D. N. & Zadeh, G. *Neuro Oncol.* **19**, 336–344 (2017).
10. Kleppe, A. *et al.* *Lancet Oncol.* **19**, 356–369 (2018).
11. Ehteshami Bejnordi, B. *et al.* *J. Am. Med. Assoc.* **318**, 2199–2210 (2017).
12. Turcan, S. *et al.* *Nature* **483**, 479–483 (2012).
13. Schwartzentruber, J. *et al.* *Nature* **482**, 226–231 (2012).
14. Wiestler, B. *et al.* *Acta Neuropathol.* **128**, 561–571 (2014).
15. Sahm, F. *et al.* *Lancet Oncol.* **18**, 682–694 (2017).
16. Korshunov, A. *et al.* *Acta Neuropathol.* **134**, 965–967 (2017).

This article was published online on 14 March 2018.

APPLIED PHYSICS

A diamond age of masers

Applications for masers — the microwave equivalent of lasers — have been hindered by their extreme operating conditions and the inability to produce continuous emissions. A diamond maser overcomes these limitations. SEE LETTER P.493

REN-BAO LIU

In 1954, scientists reported the first maser¹ — a device similar to a laser, but operating at microwave frequencies. Although lasers were not demonstrated until six years later², masers have not been as widely used as their optical counterparts. The bottleneck has been the need to operate masers under

conditions of either high vacuum or extremely low temperature (a few kelvin). On page 493, Breeze *et al.*³ present, for the first time, a maser that works continuously under ambient conditions. Such a device could lead to advances in microwave metrology and communications, and in quantum many-body physics.

The key component of a maser (or a laser) is a material known as a gain medium. In

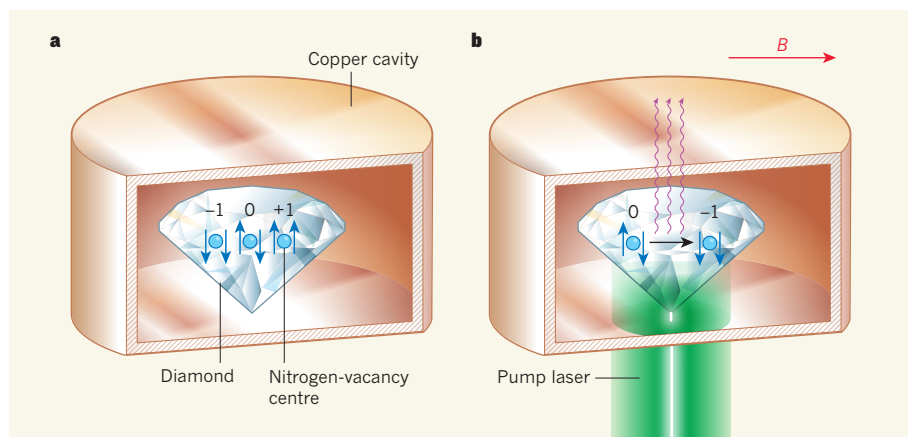


Figure 1 | A diamond maser. **a**, Breeze *et al.*³ report a microwave laser (maser) that operates continuously under ambient conditions. Their maser consists of a closed copper structure called a cavity and a millimetre-scale diamond that contains defects known as nitrogen-vacancy centres. Each of these defects has two unpaired electrons whose magnetic moments (spins; blue arrows) can point in one of two directions. Nitrogen-vacancy centres therefore have three possible spin states (denoted by -1 , 0 and $+1$). **b**, The authors applied a strong magnetic field (B) to the diamond, so that the -1 state had a lower energy than the 0 state. They then used a laser to ‘pump’ the nitrogen-vacancy centres into the 0 state. The defects relaxed to the -1 state (black arrow), and in doing so produced the microwave radiation associated with a maser.

ordinary materials, electrons usually exist in their lowest energy states and can absorb radiation by jumping to higher states. However, in a gain medium, the electron population is inverted: there are more electrons in higher states than in the lowest ones.

A photon passing through a gain medium can stimulate an electron in a higher state to jump to a lower state and emit an identical photon. If these photons are bounced back and forth between mirrors, or confined in a closed metal structure called a cavity, they can be copied many times before escaping from the system, generating a macroscopic quantum state of many identical photons. This process explains the name maser (laser): microwave (light) amplification by stimulated emission of radiation.

The different orientations of the tiny magnet (the spin) associated with an electron are known as electron spin states, and have energy separations of the right magnitude for microwave emission. However, these states are susceptible to the collisions, rotations and vibrations of atoms through an effect called spin–orbit coupling. Researchers have minimized atomic collisions in masers using high-vacuum conditions and dilute gain media, such as ammonia molecules (as in the first reported maser¹), hydrogen atoms, free electrons and rubidium gases⁴. Masers have also been built using solid-state gain media, such as ruby, and sapphire ‘doped’ with iron, in which extremely low temperatures are required to suppress the atomic vibrations⁴.

For practical applications, a solid-state maser that operates at room temperature is highly desirable. In 2012, a breakthrough towards this goal was reported in the form of a maser whose gain medium was an

organic material: a crystal of the compound *p*-terphenyl, doped with pentacene molecules⁵. Lightweight atoms (such as carbon, hydrogen and oxygen) have weak spin–orbit coupling, which means that spins in organic materials have relatively long lifetimes — they can remain in a particular energy state for a long time before jumping to a different state.

Population inversion in the *p*-terphenyl maser was achieved by ‘pumping’ the electrons in the pentacene molecules with optical radiation at a pump rate of more than 10^4 hertz. However, organic materials cannot usually withstand the intensive laser radiation required for such optical pumping — the materials often have low melting points, are evaporated by optical radiation and have poor heat conductivity. Consequently, the *p*-terphenyl maser could produce only microwave pulses, rather than continuous emission.

Building on previous work^{6–8}, Breeze and colleagues used a millimetre-scale diamond as the gain medium for their maser (Fig. 1a). Because it is composed of carbon atoms, diamond would be expected to have long spin lifetimes. The authors housed the diamond in a copper cavity and introduced free-electron spins into the diamond by adding defects called nitrogen-vacancy centres. Such a defect comprises a nitrogen atom, which replaces a carbon atom, and the void of a nearest-neighbour carbon atom. A nitrogen-vacancy centre has two unpaired electrons and three possible spin states (denoted by -1 , 0 and $+1$)⁹.

Breeze *et al.* applied a strong, uniform magnetic field to the diamond so that the -1 state had a lower energy than the 0 state. They then used a laser to pump the nitrogen-vacancy centres into the 0 state, to achieve population inversion. The defects produced

microwave radiation as they relaxed to the -1 state (Fig. 1b).

The maser required a pump power of at least 138 milliwatts. For a power of 180 mW, the authors measured the spin lifetime of the nitrogen-vacancy centres to be roughly 50 times that of the electrons in the *p*-terphenyl maser and the pump rate to be only about 300 Hz. Thanks to diamond’s high thermal conductivity (10,000 times higher than that of *p*-terphenyl), the temperature of the gain medium increased by only 35 °C when the pump power was raised to 400 mW. At room temperature, the maser worked continuously for up to 10 hours without noticeable degradation in power.

Because the frequency of microwaves produced by masers is highly stable, these devices have applications in time-keeping, high-precision spectroscopy and microwave amplification for deep-space communication and for the detection of astronomical objects. In the absence of solid-state masers that could operate at room temperature, alternative microwave sources and amplifiers were developed that were based on, for example, electronic circuits called crystal oscillators¹⁰ and sensitive detectors known as superconducting quantum-interference devices¹¹. These usually also require low temperatures. The authors’ room-temperature solid-state maser could therefore transform both microwave metrology and communications.

The good thermal conductivity of diamond and the persistence of long spin lifetimes at high temperatures¹² mean that the authors’ maser could be pumped at higher powers than they demonstrated, improving both the intensity and the stability of the emission. However, the maser’s performance is constrained by various factors, including the requirement for a strong, uniform magnetic field, temperature fluctuations caused by laser heating and low efficiency of power conversion from the pump laser to the output. Possible approaches to address these issues could include introducing the types of defect found in similar materials to diamond⁸, or borrowing ideas from other areas of research, such as superradiant lasers¹³ and lasing without population inversion¹⁴.

Finally, Breeze and colleagues’ maser could provide a platform for studying quantum many-body physics. The spins of nitrogen-vacancy centres have not only long lifetimes, but also long coherence times⁹ — the length of time for which spins can be in several different energy states at the same time, known as a quantum superposition. For this reason, nitrogen-vacancy centres have been intensively studied for quantum computing⁹ and quantum sensing¹⁵.

The interaction between many spins and many microwave photons in the authors’ maser could result in a quantum mixture that has a half-spin, half-photon nature^{16–18}. Such a mixture might offer a way to study macroscopic quantum phenomena at room

temperature. These studies could be further enriched by introducing dipole–dipole interactions between the spins, or by transforming the electron spin states into nuclear spin states.

Moreover, if the spin coherence time in the maser were longer than the photon storage time of the cavity, a photon–spin mixture could be realized in which the quantum coherence is associated mainly with the spins. The result would be a superradiant maser¹³ that, unlike the authors' maser, has an emission frequency that is insensitive to the temperature fluctuations caused by laser heating. Thanks to Breeze and colleagues, a diamond age of masers can now be envisaged. ■

Ren-Bao Liu is in the Department of Physics and the Centre for Quantum Coherence, The Chinese University of Hong Kong, Shatin, New Territories, Hong Kong, China. e-mail: rbliu@cuhk.edu.hk

- Gordon, J. P., Zeiger, H. J. & Townes, C. H. *Phys. Rev.* **95**, 282–284 (1954).
- Maiman, T. H. *Nature* **187**, 493–494 (1960).
- Breeze, J. D., Salvadori, E., Sathian, J., Alford, N. McN. & Kay, C. W. M. *Nature* **555**, 493–496 (2018).
- Bertolotti, M. *The History of the Laser* (CRC, 2004).
- Oxborrow, M., Breeze, J. D. & Alford, N. M. *Nature* **488**, 353–356 (2012).
- Jin, L. *et al. Nature Commun.* **6**, 8251 (2015).
- Poklonski, N. A. *et al. Chin. Phys. Lett.* **24**, 2088–2090 (2007).

- Kraus, H. *et al. Nature Phys.* **10**, 157–162 (2014).
- Doherty, M. W. *et al. Phys. Rep.* **528**, 1–45 (2013).
- van Beek, J. T. M. & Puiers, R. J. *Micromech. Microeng.* **22**, 013001 (2012).
- Clarke, J. & Braginski, A. I. (eds) *The SQUID Handbook: Fundamentals and Technology of SQUIDS and SQUID Systems Vol. I* (Wiley, 2004).
- Toyili, D. M. *et al. Phys. Rev. X* **2**, 031001 (2012).
- Bohnet, J. G. *et al. Nature* **484**, 78–81 (2012).
- Scully, M. O., Zhu, S.-Y. & Gavrielides, A. *Phys. Rev. Lett.* **62**, 2813–2816 (1989).
- Schirhagl, R., Chang, K., Lorez, M. & Degen, C. L. *Annu. Rev. Phys. Chem.* **65**, 83–105 (2014).
- Zhang, X., Zou, C.-L., Jiang, L. & Tang, H. X. *Phys. Rev. Lett.* **113**, 156401 (2014).
- Tabuchi, Y. *et al. Phys. Rev. Lett.* **113**, 083603 (2014).
- Rose, B. C. *et al. Phys. Rev. X* **7**, 031002 (2017).

NEURODEGENERATION

Protein aggregates caught stalling

Low-complexity protein aggregates are a hallmark of neurodegeneration. High-resolution snapshots of the structure of one such aggregate offer an unprecedented view of how these proteins disrupt crucial cellular functions.

LAURA PONTANO VAITES & J. WADE HARPER

Neurodegenerative diseases are often associated with genetic mutations that cause repetition of short sequences of nucleotides. In the disorders amyotrophic lateral sclerosis (ALS, also known as motor neuron disease) and frontotemporal dementia, such an expansion in a non-protein-coding region of the *C9orf72* gene^{1,2}, leads to aberrant translation products that contain repetitive stretches of glycine and alanine amino-acid residues. These 'poly(GA)' products form aggregates in neurons, and have been implicated in the disruption of a key cellular process in which complexes called proteasomes degrade proteins^{3,4}. However, the biochemical basis for this disruption, and how it might promote disease, is poorly understood. Writing in *Cell*, Guo *et al.*⁵ precisely map the organizational and structural features of poly(GA) aggregates and associated macromolecular complexes in neurons using a technique called 3D cryo-electron tomography (cryo-ET), to provide direct visualization of how proteasomes are disrupted by poly(GA) proteins.

Cryo-ET in 3D uses electron microscopy to view very thin, frozen but hydrated sections of a cell from various angles. The resulting images are combined to produce a 3D image called a tomogram. Guo *et al.* used 3D cryo-ET to visualize neurons that had been genetically engineered to express a poly(GA) tract that contained either 175 or 73 repeats. The tracts

were fused with a green fluorescent protein that enabled their precise position to be determined using correlative light microscopy. The engineered protein mimics poly(GA) tracts that are produced from *C9orf72* expansion, which take a long time to form *in vivo*. The authors found that poly(GA) proteins

form highly clustered and often bifurcated twisted ribbon structures that are of relatively uniform thickness, but of variable length and width, similar to poly(GA) structures previously observed by conventional electron microscopy *in vitro*⁶.

The value of the authors' work lies not only in their observation of the structure of poly(GA) aggregates in detail in cells, but in their comparison of poly(GA) with aggregates formed through a different genetic expansion — a glutamine repeat (poly(Q)) tract, which causes the neurodegenerative disorder Huntington's disease, and which the same group analysed by 3D cryo-ET last year⁷. This comparison revealed structural differences that could explain dissimilarities in pathogenic mechanisms between the conditions.

First, the aggregates formed in each case are themselves structurally distinct. Poly(Q) proteins form fibril structures that show little

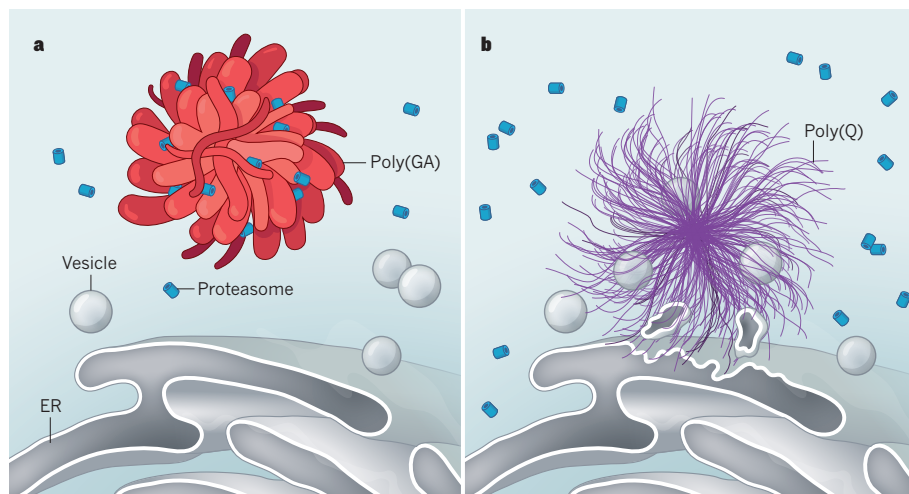


Figure 1 | Contrasting mechanisms of aggregate toxicity. **a**, In some cases of the neurodegenerative disorder amyotrophic lateral sclerosis, long chains of glycine and alanine amino-acid residues (dubbed poly(GA) tracts) aggregate in neurons. Guo *et al.*⁵ show, through high-resolution structures in cells, that poly(GA) tracts form ribbon-like aggregates that capture protein complexes called proteasomes, which normally process other proteins for degradation. Such capture causes proteasome stalling, providing an explanation for the toxicity of this aggregate. Poly(GA) aggregates do not bind membrane-bound organelles such as vesicles and the endoplasmic reticulum (ER). **b**, By contrast, repetitive tracts of the amino acid glutamine (poly(Q) tracts), which are associated with Huntington's disease, form fibril-like aggregates⁷. These aggregates deform the membranes of vesicles and the ER, suggesting that different aggregates cause neurodegeneration through different mechanisms.



50 Years Ago

The British Home Office seems to be modifying slightly its attitude to the tests by which motorists in Britain can now be convicted of driving under the influence of alcohol. A recent paper ... by Professor J. B. Payne ... suggests that the methods used ... are far from accurate ... So far the Home Office has not been forced to act, because no motorist accused of driving under the influence of drink has quoted Professor Payne's work in his defence ... Originally ... police surgeons were advised to take small samples of capillary blood for use in the test, although it was also open to them to take venous blood if they preferred. The work at the Royal College of Surgeons suggests that the latter is likely to give more accurate results ... The Home Office has now sent a circular to police authorities pointing out that it is within their discretion to take venous rather than capillary blood ... The circular also points out that the motorist has the right to keep a sample of his own blood for independent analysis.

From *Nature* 23 March 1968

100 Years Ago

Now that our rations of food, particularly of meat and wheat bread, have been so appreciably reduced the necessity of arranging our diet so as to ensure a sufficient supply of those elusive substances, the so-called "vitamines", is more important than ever. It is known that these substances exist in certain foods, and that an adequate supply of them is necessary to health, but they have not yet been isolated in a pure condition, although several workers claim to have done so successfully. As a result of some recent work, McCollum and Davis concluded that two distinct types of vitamine exist, the "fat-soluble A" and the "water-soluble B".

From *Nature* 21 March 1918

branching and are less densely packed than poly(GA) ribbons⁷.

Second, when the authors used powerful computational approaches to search for known macromolecular complexes in each aggregate, they found many proteasomes incorporated in poly(GA) aggregates (Fig. 1). Indeed, biochemical data suggested that as many as 50% of the proteasome complexes in the neuron become highly entangled within poly(GA) ribbons. Removal of proteasomes from their normal location in cells through this sequestration mechanism might explain the reduced proteasomal activity in cells harbouring these aggregates^{4,8}. Complexes called ribosomes, which mediate protein production and are comparable in size to proteasomes, were largely excluded from poly(GA) ribbons, suggesting that poly(GA) aggregates are actively recruited or retained by proteasomes. By contrast, poly(Q) fibrils did not contain proteasomes, but formed close contacts with membranes from multiple types of organelle. This interaction leads to deformation of the membranes around organelles, such as the endoplasmic reticulum. Such deformation might alter pathways involved in protein translation, trafficking and degradation⁷.

The proteasome consists of a barrel-shaped core particle in which substrate cleavage takes place, and one or two regulatory particles that cap the ends of the barrel, restricting access to the core so that only proteins tagged with the pro-degradation molecule ubiquitin can enter. Regulatory particles have been observed in multiple conformations^{9,10}, indicating that proteasomes progress through a reaction cycle that involves ground, committed and substrate-engaged states. Guo *et al.* used computational particle averaging to quantify the proteasomal states (technique reviewed in ref. 11), and found both ground and substrate-engaged states within poly(GA) aggregates. They also found a large increase in the proportion of doubly capped proteasomes (indicating engagement with substrate) compared with control neurons that did not contain poly(GA) products. Almost one-quarter of the proteasomes within aggregates adopted a conformation recently described⁹ as substrate-engaged yet stalled, in which the substrate becomes trapped in the barrel. And that proportion rose to 36% for those proteasomes closest to poly(GA) ribbons.

Why might this stalling occur? The authors' tomographic reconstructions revealed numerous regions of electron density located between a poly(GA) ribbon and the site where the protein RAD23 binds to the proteasome. RAD23 is involved in recruiting ubiquitin-tagged proteins to the proteasome and is known to be enriched in poly(GA) aggregates⁸. Thus, this electron density could indicate RAD23-associated ubiquitin that is attached to proteins within the aggregate. Which protein or proteins the proteasome is choking on is currently

unclear, although possibilities include the poly(GA) peptides themselves, which probably inhibit proteasome activity directly. Regardless of the mechanism, it seems likely that depletion of proteasomal activity in the cell proper would be detrimental to protein-degradation pathways, thereby contributing to cellular toxicity.

These results raise several important questions. First, pathogenesis in cases of ALS driven by *C9orf72* expansion has been linked both to changes mediated by poly(GA) formation and to changes caused by reduced production of *C9orf72* protein — but what are the relative contributions of each mechanism? *C9orf72* is part of a complex involved in autophagy¹², a process by which cellular material, including proteins, is degraded and recycled. It is therefore possible that reduced *C9orf72* levels conspire with poly(GA)-dependent proteasome inhibition to increase neuronal toxicity. Second, is toxicity promoted by the capture of other proteins within poly(GA) aggregates? One candidate is the autophagy cargo receptor p62, which is known to accumulate in poly(GA) aggregates⁸. Third, several molecular machines involved in disassembling aggregates do not accumulate in poly(GA) structures, but the reasons for this are unclear.

Finally, although poly(GA) is the most abundant repetitive protein produced by *C9orf72* expansion, it is not the only one — mutation can also produce tracts of glycine-arginine (poly(GR)) and proline-arginine (poly(PR)). How do the structures of these other aggregates compare to that of poly(GA) proteins? Most data on poly(GR) and poly(PR) aggregates indicate that they do not accumulate proteasomes, suggesting alternative toxicity mechanisms^{13,14}. Further analysis by 3D cryo-ET, and analysis of natural products of *C9orf72* expansion rather than the engineered product used in the current study, might clarify the similarities and differences between the aggregates that occur in patients and the model aggregate structures studied by Guo and colleagues.

In sum, the current work highlights the unprecedented resolution of 3D cryo-ET for visualizing fundamental processes within cells¹¹. Moreover, it sets the stage for a more comprehensive mechanistic understanding of aggregate-associated neurodegenerative diseases. ■

Laura Pontano Vaites and J. Wade Harper are in the Department of Cell Biology, Harvard Medical School, Boston, Massachusetts 02115, USA.

e-mail: wade_harper@hms.harvard.edu

- DeJesus-Hernandez, M. *et al.* *Neuron* **72**, 245–256 (2011).
- Taylor, J. P., Brown, R. H. Jr & Cleveland, D. W. *Nature* **539**, 197–206 (2016).
- Yamakawa, M. *et al.* *Hum. Mol. Genet.* **24**, 1630–1645 (2015).
- Zhang, Y.-J. *et al.* *Acta Neuropathol.* **128**, 505–524 (2014).

5. Guo, Q. *et al. Cell* **172**, 696–705 (2018).
6. Chang, Y.-J., Jeng, U.-S., Chiang, Y.-L., Hwang, I.-S. & Chen, Y.-R. *J. Biol. Chem.* **291**, 4903–4911 (2016).
7. Bauerlein, F. J. B. *et al. Cell* **171**, 179–187 (2017).
8. May, S. *et al. Acta Neuropathol.* **128**, 485–503 (2014).

9. Wehmer, M. *et al. Proc. Natl Acad. Sci. USA* **114**, 1305–1310 (2017).
10. Chen, S. *et al. Proc. Natl Acad. Sci. USA* **113**, 12991–12996 (2016).
11. Beck, M. & Baumeister, W. *Trends Cell Biol.* **26**, 825–837 (2016).

12. Amick, J. & Ferguson, S. M. *Traffic* **18**, 267–276 (2017).
13. Lin, Y. *et al. Cell* **167**, 789–802 (2016).
14. Lee, K.-H. *et al. Cell* **167**, 774–788 (2016).

This article was published online on 19 March 2018.

SOLAR SYSTEM

Calcium signals in planetary embryos

The calcium–isotope composition of planetary bodies in the inner Solar System correlates with the masses of such objects. This finding could have implications for our understanding of how the Solar System formed. [SEE LETTER P.507](#)

ALESSANDRO MORBIDELLI

For decades, researchers have debated how planetary bodies in the Solar System with diameters of more than a few hundred kilometres were formed. Current models¹ suggest that such bodies accumulated (accreted) small particles known as pebbles that were drifting towards the Sun in a cloud of gas and dust called the protoplanetary disk. However, these models have difficulty in explaining the diverse composition of objects in the Solar System: if all such bodies grew by accretion from the same flow of pebbles, then why do they have different compositions? On page 507, Schiller *et al.*² provide a possible answer to this question, on the basis of their observation that the composition of calcium isotopes in planetary bodies correlates with the masses of these bodies.

The authors measured the calcium–isotope content of samples from the parent bodies of types of meteorite known as angrites and ureilites, as well as from Earth, Mars and the asteroid Vesta. They present their data in terms of the value $\mu^{48}\text{Ca}$, which expresses the ratio of two calcium isotopes (^{48}Ca and ^{44}Ca) in a sample relative to that in a terrestrial reference standard, and is given as parts per million (p.p.m.). The authors identified a positive correlation between $\mu^{48}\text{Ca}$ and planetary-body mass for Earth, Mars and Vesta, which have known masses, and for the parent bodies of angrites and ureilites, the masses of which were inferred from thermal models^{3,4}.

Schiller *et al.* account for this correlation by assuming that the pebble-accretion scenario is correct, but by rejecting the conventional idea that bodies of different masses grew at different rates throughout the lifetime of the protoplanetary disk. Instead, the authors suggest that all bodies grew at the same rate, but stopped growing at different times: smaller bodies ceased accretion earlier than did larger ones.

This unorthodox view of growth is, in fact, supported by numerical simulations⁵, which show that growing bodies perturb each other's orbits around the Sun. Bodies that acquire eccentric or inclined orbits

stop accumulating pebbles, whereas those that remain on circular trajectories in the mid-plane region of the disk continue to grow. The correlation of $\mu^{48}\text{Ca}$ values with planetary-body masses therefore also becomes a correlation with the timescale of the growth of such bodies. The authors confirm this using growth timescales for angrites, ureilites, Vesta and Mars that were inferred by nuclear chronometry (a dating technique that uses the decay of radioactive isotopes) and thermal models.

Schiller and colleagues propose that material in the inner part of the disk initially had low $\mu^{48}\text{Ca}$ values (about –150 p.p.m.). Planetary bodies grew from this matter until they reached the size of the ureilite parent body (about 200 km in diameter). The inner disk then started to be fed with pebbles that drifted

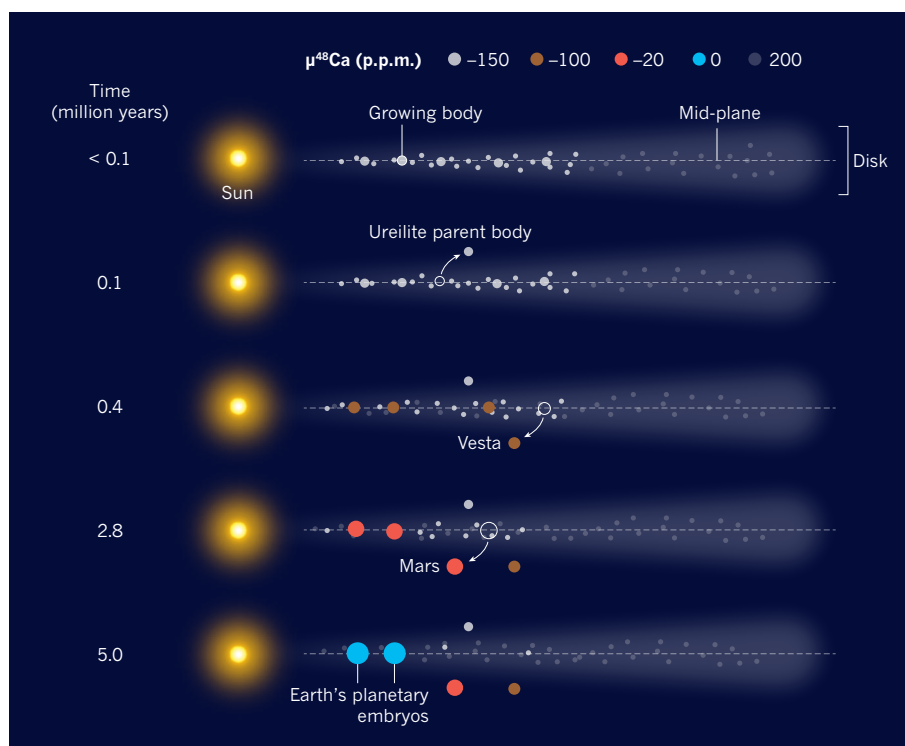


Figure 1 | A model of planetary accretion. Schiller *et al.*² report a correlation between the calcium–isotope contents and the masses of planetary bodies in the inner Solar System, and propose a model to explain this observation. They present their data in terms of the value $\mu^{48}\text{Ca}$, which expresses the ratio of two calcium isotopes (^{48}Ca and ^{44}Ca) in a body relative to that in a terrestrial reference standard, and is given as parts per million (p.p.m.). In the authors' model, the disk of gas and dust that surrounded the newly formed Sun initially contained inner-disk material with low $\mu^{48}\text{Ca}$ values (–150 p.p.m.) and outer-disk particles with high $\mu^{48}\text{Ca}$ values (200 p.p.m.). The planetary bodies grew by accumulating (accreting) matter from the inner disk. Over time, the inner-disk material was replaced by particles that flowed in from the outer disk, which progressively increased the mean $\mu^{48}\text{Ca}$ value of the inner disk. Bodies continued to grow until they were displaced (curved arrows) from the mid-plane of the disk by perturbations from other bodies. The times indicate stages in the disk's lifetime; in particular, the stages at which various bodies were formed — the parent bodies of a type of meteorite known as ureilites, the asteroid Vesta, Mars and the planetary embryos that eventually produced Earth.

in from the outer part of the disk. Such pebbles had high $\mu^{48}\text{Ca}$ values (about 200 p.p.m.), which are typical of the most primitive meteorites (carbonaceous chondrites) that are thought to have formed beyond Jupiter⁶. The mean $\mu^{48}\text{Ca}$ value in the inner disk therefore increased progressively. As a consequence, bodies that continued to grow to the size of Vesta (530 km in diameter) reached $\mu^{48}\text{Ca}$ values of ~100 p.p.m., and those that grew to the size of Mars (6,800 km in diameter) reached $\mu^{48}\text{Ca}$ values of ~20 p.p.m (Fig. 1).

Earth and the Moon have $\mu^{48}\text{Ca}$ values of about 0 p.p.m. It is generally assumed that Earth formed after the disappearance of the protoplanetary disk, as a result of collisions of planetary embryos with masses similar to that of Mars; the body that collided with early Earth to give rise to the Moon was one of these embryos. However, this formation scenario is not possible if Schiller and colleagues' proposal is correct. Planetary embryos with the mass of Mars would have $\mu^{48}\text{Ca}$ values of ~20 p.p.m., and therefore an 'Earth' resulting from the merger of these bodies would have a similar composition.

Instead, for Earth to have a $\mu^{48}\text{Ca}$ value of about 0 p.p.m., the embryos would need to have grown by pebble accretion to masses roughly half that of Earth. In turn, the Moon-forming impactor would need to have had a mass comparable to that of the embryos to have the same $\mu^{48}\text{Ca}$ value, supporting the theory that the Moon resulted from the collision of two bodies with half-Earth masses⁷.

A potential problem is that hafnium–tungsten radioactive chronometry indicates that Earth reached 63% of its present mass only after a duration of between 11 million and 24 million years, depending on the type of core–mantle equilibration that occurred during the collision of the embryos⁸. It is therefore difficult to imagine that such embryos grew to bodies with half-Earth masses in the disk's putative lifetime of 5 million years⁹.

But perhaps the embryos did reach the size of one-third of Earth's mass (three times the mass of Mars). Such a proposal cannot be discounted as a compromise between the authors' correlation of calcium–isotope compositions with planetary masses and the chronological constraints on Earth's accretion. The fraction of Earth's mass that would have been accreted from the outer disk (estimated at about 40% in the authors' study) would be higher than calculated previously^{10,11}, but such computations did not consider some of the building blocks of Earth to have ureilite-like compositions.

Schiller and colleagues' view of accretion in the Solar System is in sharp contrast with that presented by two previous studies^{6,12}. These reports concluded that the flux of pebbles from the outer disk was shut down during the first million years of the disk's lifetime by the formation of Jupiter. This prevented bodies in the inner Solar System from accumulating

large amounts of water ice, explaining why such bodies are mostly dry¹², and maintained an isotopic dichotomy between two types of meteorite: ordinary and carbonaceous chondrites⁶.

The composition of ordinary chondrites is difficult to account for using Schiller and colleagues' model. Unlike large bodies, which should grow continuously by pebble accretion, the smaller parent bodies of chondrites should form suddenly, from clusters of pebbles that are generated by a mechanism known as a streaming instability¹. In the case of ordinary chondrites, this would have happened in the inner Solar System at a late stage in its formation, after the time by which Mars had accreted most of its mass. Ordinary chondrites should therefore represent a snapshot of the composition of the late disk, and have a positive value of $\mu^{48}\text{Ca}$ in the authors' model. But, in reality, such chondrites have $\mu^{48}\text{Ca}$ values of ~35 p.p.m.

Schiller *et al.* explain this discrepancy by speculating that pebble accretion and the streaming instability have differing preferences for pebble size. The pebbles that came from the outer disk were small and, although they were efficiently accreted by large bodies, they barely participated in the streaming instability. Consequently, ordinary chondrites formed mostly from larger pre-existing pebbles called chondrules, which typically have negative values of $\mu^{48}\text{Ca}$. The validity of this proposal

will need to be checked using high-resolution numerical simulations of the pebble-accretion and streaming-instability processes.

The authors' work adds a missing piece to the jigsaw puzzle of planet formation that will need to be connected with the other pieces provided by isotopic, chemical, chronological and dynamical constraints. Although the puzzle seems more complete than before, perhaps some other key pieces are still missing. ■

Alessandro Morbidelli is at the *Observatoire de la Côte d'Azur, CS 34229, 06304 Nice Cedex 4, France.*

e-mail: morby@obs-nice.fr

1. Johansen, A., Mac Low, M.-M., Lacerda, P. & Bizzarro, M. *Sci. Adv.* **1**, e1500109 (2015).
2. Schiller, M., Bizzarro, M. & Fernandes, V. A. *Nature* **555**, 507–510 (2018).
3. Schiller, M., Baker, J. A. & Bizzarro, M. *Geochim. Cosmochim. Acta* **74**, 4844–4864 (2010).
4. Wilson, L., Goodrich, C. A. & Van Orman, J. A. *Geochim. Cosmochim. Acta* **72**, 6154–6176 (2008).
5. Levison, H. F., Kretke, K. A. & Duncan, M. J. *Nature* **524**, 322–324 (2015).
6. Kruijer, T. S., Burkhardt, C., Budde, G. & Kleine, T. *Proc. Natl Acad. Sci. USA* **114**, 6712–6716 (2017).
7. Canup, R. M. *Science* **338**, 1052–1055 (2012).
8. Kleine, T. *et al. Geochim. Cosmochim. Acta* **73**, 5150–5188 (2009).
9. Mamajek, E. E. *AIP Conf. Proc.* **1158**, 3–10 (2009).
10. Fitoussi, C., Bourdon, B. & Wang, X. *Earth Planet. Sci. Lett.* **434**, 151–160 (2016).
11. Dauphas, N. *Nature* **541**, 521–524 (2017).
12. Morbidelli, A. *et al. Icarus* **267**, 368–376 (2016).

NEURODEVELOPMENT

A mixed model of neuronal diversity

Two groups have sequenced RNA from thousands of single cells, making the deepest ventures yet into the origins of neuronal diversity in the neocortex of the developing mammalian brain. [SEE ARTICLE P.457](#)

**LUDOVIC TELLEY
& DENIS JABAUDON**

Biologists have long been preoccupied with classifying diverse objects into meaningful categories. Even at the beginning of genetics, Gregor Mendel classified peas as smooth or wrinkled, to infer the mechanisms underlying the inheritance of specific features. Neurons, like peas, come in different shapes and types, but how many neuronal cell types exist and how neuronal diversity emerges during development is unclear. Answering these questions is important, because the diversity of neurons determines the diversity of circuits that can be built and, by extension, the scope of an animal's

behavioural repertoire. Two groups (one writing in *Nature*¹ and one in *Science*²) now investigate the origins of neuronal diversity in the developing neocortex, a brain region that is the source of mammals' complex behavioural and cognitive abilities³.

Two opposing — although not mutually exclusive — scenarios account for the generation of distinct kinds of neuron across the nervous system, and in the cerebral cortex in particular^{3–6}. In the first scenario, diverse neuronal subtypes are born from correspondingly diverse progenitor cells (Fig. 1a). This is known as the premitotic model, because diversity arises in progenitors, before neurons are actually generated (neuronal birth occurs through mitotic cell division). In this

scenario, the diversity of progenitors might result from the cells' locations in different parts of the brain, or from their generation at different times. In the second scenario, known as the postmitotic model, progenitors are homogeneous, and diversity arises in neurons as they develop, including through interactions with the surrounding environment, which sculpt these cells into more-refined adult cell types (Fig. 1b).

In the current studies, the groups investigated how diversity arises in the neocortex by using single-cell RNA sequencing to identify and define distinct cell types by their transcriptional signatures. In contrast to previous approaches, in which cells were lumped together to yield average transcriptional activities, this technology allows several thousand individual cells to be simultaneously singled out and sequenced, yielding a more fine-grained view of neurons' molecular identities⁷.

In the first study, Mayer *et al.*¹ (page 457) focused on a population of neurons in the mouse neocortex that releases the inhibitory neurotransmitter molecule GABA. These neurons arise from progenitors located in three transient structures of the developing mammalian brain, called ganglionic eminences⁸. Using single-cell RNA sequencing, the authors revealed that transcriptional programs are largely conserved in progenitors across the ganglionic eminences. Only a select set of genes is differentially expressed between the three regions — a limited level of premitotic diversity that is consistent with the postmitotic model.

Next, the researchers used elaborate bioinformatics approaches to further reconstruct the developmental trajectory of each neuronal subtype. This analysis demonstrated a link between the initial diversity of immature neurons and the diversity in mature neuronal subtypes, consistent with the premitotic model. The group thus showed that mature neuronal properties are already distinguishable in a rough form in newborn neurons. Subtypes then become more crisply defined as neurons, poised in the genetic ground state dictated by progenitors; they then differentiate and interact with the environment. Altogether, these data support a mixed model of development in which diversification occurs in both pre- and postmitotic cells (Fig. 1c).

In the second study, Nowakowski *et al.* focused on excitatory neurons that produce the neurotransmitter glutamate, in two neocortical

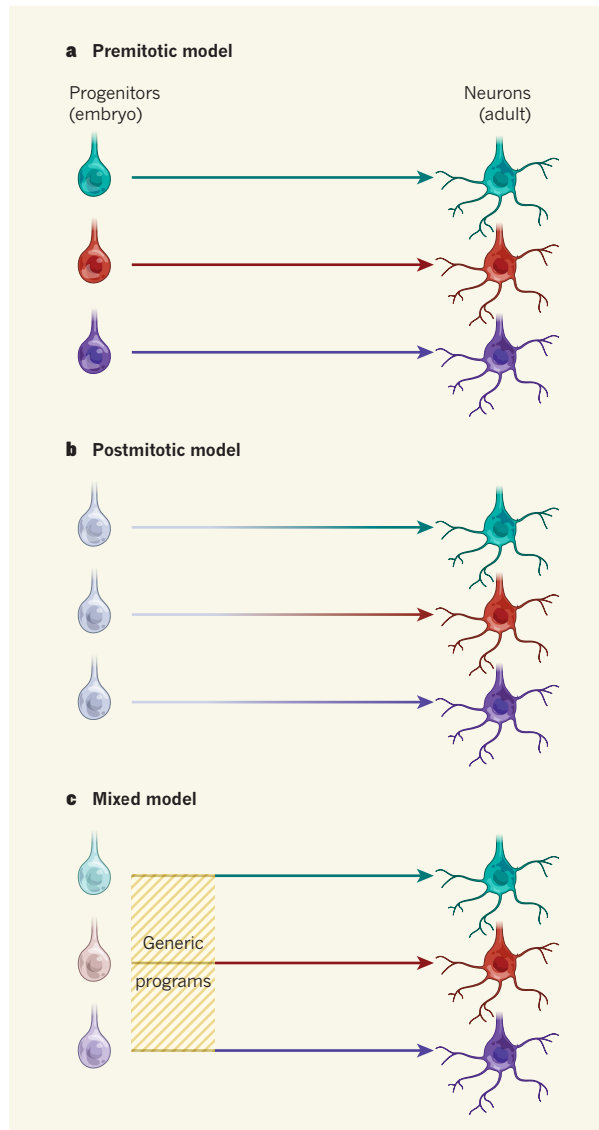


Figure 1 | Sources of neuronal diversity. **a**, In the 'premitotic' model, neuronal diversity arises in progenitor cells, owing to differences in either their location or the timing of their generation (different colours represent different gene-expression patterns). Diverse progenitors give birth to diverse neurons (the beginning of each arrow indicates neuron birth), which mature into diverse adult neurons. **b**, In the 'postmitotic' model, diversity emerges only after progenitors have differentiated into neurons, owing in part to interactions with the environment. **c**, Two groups^{1,2} provide evidence that, in the neocortex of the mammalian brain, relatively few differences exist across progenitors (similar to the postmitotic model), but these differences are sufficient to drive neuronal diversity (as in the premitotic model), generating a mixed model. There seem to be generic transcriptional programs across neuronal types in the early stages of their development. (Modified from ref. 5.)

areas in human fetuses — the prefrontal cortex and the primary visual cortex, which are involved in behavioural planning and in vision, respectively. Neurons in these regions are organized into archetypal cortical columns, which are core functional units of circuits. The precise structure of these columns is thought to be tweaked across cortical areas to allow for different functions^{3–5}. How these spatial differences emerge is poorly understood,

but explanations involving pre- or postmitotic divergences have both been put forward^{4,5}.

The authors explored how these spatial differences emerge using similar experimental and analytical tools to those of Mayer and colleagues. Nowakowski *et al.* found that only select sets of genes are differentially expressed between progenitors across the two cortical regions, and that differences increase as neurons mature. These findings thus support a 'soft' premitotic situation that leaves leeway for extrinsic factors to drive the developmental trajectories of postmitotic newborn neurons (Fig. 1c). Despite fundamental differences in the biology of the cell types studied, then, the groups reach similar conclusions.

An intriguing finding made by Nowakowski *et al.* is that glutamate-producing neurons share their initial transcriptional trajectories with GABA-producing neurons. Similarly, Mayer *et al.* found that newborn neurons from all ganglionic eminences initially transit through overlapping transcriptional ground states. Thus, neuronal subtypes seem to emerge from subtype-specific processes that are superimposed on more-generic developmental programs.

These two studies shed long-awaited light on the origin of neuronal subtypes. However, the quest to understand neuronal diversity is only just beginning. Mendel's classification of peas was later revisited by Raphael Weldon, who argued that there were many more categories than originally described^{9,10}. Similarly, other studies might reveal more diversity than is reported in the current papers because, inevitably, the current studies are constrained by the criteria used to define cell classes.

To explain, single-cell RNA sequencing, in principle, provides an objective and comprehensive criterion for cellular classification. However, the granularity of the transcriptional assessment — factors such as sequencing quality and which kinds of RNA are analysed — is a key parameter in delineating cell types. Thus, future

studies might reveal more progenitor types than those found here, and tilt the balance towards a more premitotic view of diversity. In fact, one recently published study¹¹ reported substantial diversity in immature neurons. Furthermore, sequencing approaches differentiate between cell types only on the basis of molecular differences; other criteria such as electrophysiological properties and anatomy are probably crucial, even at

early developmental stages, for a complete definition of cell types.

How identity progresses with time in postmitotic neurons remains elusive. It will be particularly useful to dissect the extent to which external factors, including sensory signals from the peripheral nervous system, drive differentiation and further diversification of neurons within these genetically poised subtypes. Resolving these issues will be central not only to addressing the mechanisms that generate neuronal diversity, but also to characterizing and understanding inter-individual differences in circuit structure and function.

Finally, both studies identify disease-related

genes among the cell-type-specific transcripts, including some associated with autism and schizophrenia. As such, they provide a valuable resource for those seeking to understand the mechanisms underlying these disorders. ■

Ludovic Telley is in the Department of Fundamental Neuroscience, University of Lausanne, 1005 Lausanne, Switzerland.

Denis Jabaudon is in the Department of Basic Neurosciences, University of Geneva, 1211 Geneva, Switzerland.

e-mails: ludovic.telley@unil.ch;
denis.jabaudon@unige.ch

1. Mayer, C. *et al. Nature* **555**, 457–462 (2018).
2. Nowakowski, T. J. *et al. Science* **358**, 1318–1323 (2017).
3. Jabaudon, D. *Nature Commun.* **8**, 16042 (2017).
4. Rakic, P. *Science* **241**, 170–176 (1988).
5. O'Leary, D. D. M. *Trends Neurosci.* **12**, 400–406 (1989).
6. Puelles, L. & Rubenstein, J. L. R. *Trends Neurosci.* **26**, 469–476 (2003).
7. Johnson, M. B. & Walsh, C. A. *Curr. Opin. Neurobiol.* **42**, 9–16 (2017).
8. Wamsley, B. & Fishell, G. *Nature Rev. Neurosci.* **18**, 299–309 (2017).
9. Radick, G. *Nature* **533**, 293 (2016).
10. Weldon, W. F. R. *Biometrika* **2**, 44–55 (1902).
11. Mi, D. *et al. Science* <http://dx.doi.org/10.1126/science.aar6821> (2018).

This article was published online on 5 March 2018.

PLANT BIOLOGY

A cellular passage to the root interior

Water-conducting tissues inside plant roots are surrounded by impermeable cells. This protective barrier is punctured by 'passage cells', which are thought to regulate nutrient uptake. How these cells form has now been revealed. [SEE LETTER P.529](#)

SEDEER EL-SHOWK & ARI PEKKA MÄHÖNEN

Plants need to take up water and nutrients through their roots, while keeping out pathogens and toxins. To achieve this, the roots' inner transport tissues are enclosed in a protective, impermeable barrier of endodermal cells, interspersed with 'openings' in the form of a specific type of endodermal cell called a passage cell, which is thought to serve as a cellular gatekeeper, controlling access to the root interior^{1–4}. On page 529, Andersen *et al.*⁵ describe molecular mechanisms that control the formation of passage cells, and show how the numbers of such cells are regulated by nutrient availability. The findings offer insights into the formation of these key plant cells, and link this to the patterning processes that form the embryonic root, providing an intriguing demonstration of the continuity of developmental mechanisms.

As endodermal cells develop, the impermeable polymers lignin and suberin are deposited in the cell walls. The deposits build up as the cells mature, forming an impermeable barrier⁴ that isolates the adjacent inner tissues from the root's outer cell layers and the soil. Once this process is complete, it is thought that nutrients and water can reach the interior only by means of passage cells⁴. These cells contain lignin deposited in an arrangement called the Casparian strip in their cell walls, but lack suberin, and thus offer a permeable route for molecular transport^{4,5}. The extent of 'suberization' of endodermal cells is regulated by the

levels of the hormones abscisic acid (ABA) and ethylene, and can be reversibly affected by both nutrient availability and stress⁶. The extent of suberization is thought to be a key factor in the rate of nutrient uptake⁶. Although the role and regulation of suberization in endodermal cells are understood, the molecular mechanisms controlling the development of passage cells have remained a mystery.

Passage cells have been observed in many plants species^{1–3}, and are formed in a region of the root above the growing root tip. Andersen

“Plants can dynamically adjust the number of passage cells according to nutrient status.”

et al. used the plant *Arabidopsis thaliana* as their model system. To confirm that passage cells form in its roots, they tracked the expression of a gene required for suberin synthesis by monitoring a fluorescent marker protein. They identified suberin-deficient passage cells distributed at seemingly random locations along the root's length. The cells were consistently positioned around the root's circumference near the developing xylem, one of two types of water- and nutrient-conducting tissue (Fig. 1).

To investigate the nature of the link between xylem and passage cells, the authors tested plants in which genetic mutations altered the pattern of xylem development. Two of the mutant plants^{7,8} had fewer passage cells than normal. These plants had defects in the genes

AHP6 or *LOG4*, which encode proteins that act respectively in the signalling and biosynthesis of the hormone cytokinin. Cytokinin affects the transport and signalling of the hormone auxin to regulate the development and patterning of water-conducting tissues^{8,9}. Auxin, in turn, drives the expression of *AHP6*, which encodes a protein that inhibits cytokinin signalling in the developing xylem^{7,9}. The authors found that *AHP6* protein (fused to a fluorescent marker protein) diffused from the developing xylem into the adjacent endodermal cells, probably through cell-connecting nanostructures called plasmodesmata, and that this process was required for passage-cell formation.

The finding led Andersen and colleagues to suspect that a feedback loop between cytokinin and auxin signalling might be involved in passage-cell formation. They therefore analysed plants expressing marker proteins to monitor the levels of cytokinin signalling. They found that passage cells have low levels of cytokinin signalling, or lack it altogether, unlike their neighbouring suberized endodermal cells. Analysis of another marker protein showing the level of auxin signalling revealed an auxin response in all of the endodermal cells near the developing xylem, including passage cells. Seedlings grown with auxin had more passage cells than did plants that did not receive an auxin boost. Conversely, seedlings treated with cytokinin contained fewer passage cells than untreated seedlings. Opposing roles for auxin and cytokinin are a ubiquitous theme in plant development, and the presence of such a feedback loop links passage-cell formation to patterning processes acting in the transport tissues and elsewhere.

The authors investigated whether auxin and cytokinin might affect ABA-mediated suberin deposition. Plants that were engineered to express cytokinin-signalling inhibitors throughout the endodermis had low suberin deposition in all endodermal cells, and this outcome was unaffected by the addition of ABA. However, it is unclear why ABA-mediated suberin deposition requires cytokinin signalling. Plants were engineered to express auxin-signalling inhibitors throughout the

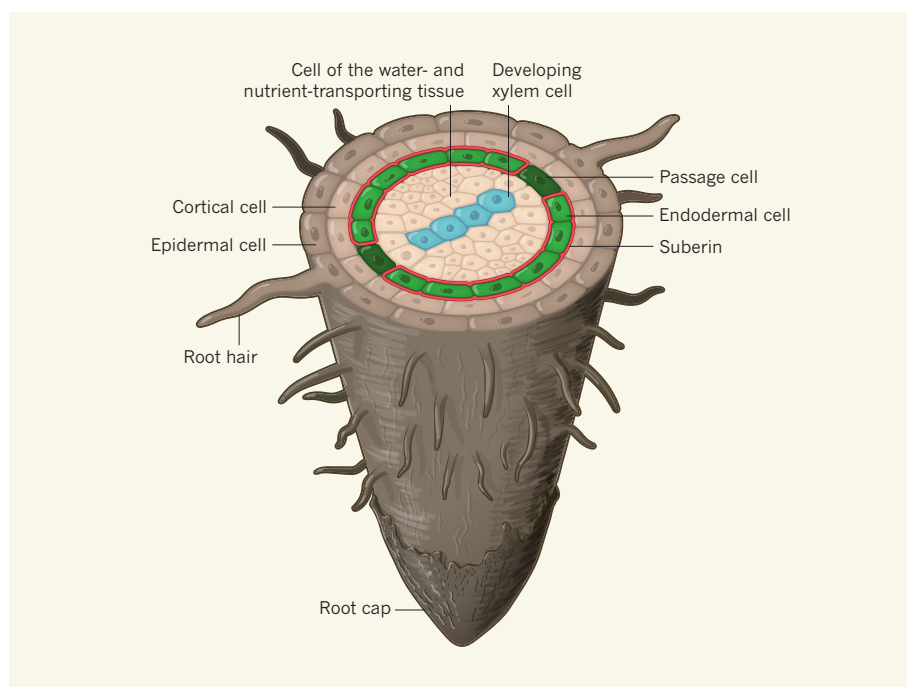


Figure 1 | Passage cells in the plant root. A protective layer of endodermal cells, rendered impermeable by deposition of the polymer suberin (red), surrounds the water- and nutrient-transport tissues at the core of the plant root. But certain endodermal cells, known as passage cells, do not contain suberin, and these cells have been proposed as cellular gatekeepers that control how water and nutrients reach the inner root tissues^{1–4}. Andersen *et al.*⁵ report studies in the model plant *Arabidopsis thaliana* that shed light on how passage cells develop. The authors find that the cells form adjacent to the region of developing xylem (blue), which is part of the plant's transport tissues. The protein AHP6 (not shown) moves from the developing xylem into endodermal cells. It inhibits the signalling pathway mediated by the hormone cytokinin, and prevents suberin from being deposited in the cell walls in response to the hormone abscisic acid (not shown). The lack of suberin enables the formation of passage cells.

endodermis, the number of passage cells was reduced compared with controls, and ABA addition had no effect on this, either.

Andersen and colleagues propose that most endodermal cells usually have a high level of cytokinin signalling activity, enabling them to undergo ABA-mediated suberization, and that these cells eventually reach a level of suberization that makes them impermeable. However, endodermal cells adjacent to the developing xylem receive AHP6, resulting in low levels of cytokinin signalling, and these cells also have high auxin signalling. These characteristics render some of the cells unresponsive to ABA, so they remain unsuberized and form passage cells. The authors suggest that stochastic differences in cytokinin-mediated ABA sensitivity may determine which of the cells become passage cells, but it will be fascinating to discover whether this process is fine-tuned by other factors, such as nutrient sensing.

Additional work will be required to elucidate the genetic and hormonal interactions that mediate this patterning network. Does cytokinin regulate sensitivity to ABA in the endodermis, and do auxin transporter proteins concentrate auxin to aid passage-cell formation? Alternatively, is auxin distribution in the endodermis determined by the auxin dynamics in the underlying transport tissues¹⁰?

Andersen and colleagues also investigated passage-cell function. Although it has been proposed that these cells are involved in nutrient intake, only a few studies have investigated their role^{4,5}. A previous study¹¹ in *A. thaliana* roots found that the expression of a gene encoding a phosphate transporter protein is correlated with the location of passage cells. Andersen and colleagues confirmed this finding, and showed that the transporter gene and several related genes are expressed in adjacent cortical and epidermal cells in the outer layers of the root. This pattern of phosphate transporters suggests that phosphate, and perhaps other nutrients, might be funnelled through passage cells towards the water- and nutrient-transport system.

The authors found that nutrient deficiency leads to more-widespread expression of the phosphate transporter in xylem-associated endodermis and a reduction in suberization of these cells. This shows that plants can dynamically adjust the number of passage cells according to a plant's nutrient status. Finally, Andersen *et al.* discovered that monitoring the expression of phosphate transporters allowed passage cells to be identified at an early stage of development. Identification occurred before the initial deposition of suberin in neighbouring endodermal cells that do not form passage

cells, providing a tool for future studies of the early events in passage-cell formation.

This latest work opens the door to further investigation of the role of passage cells. Genetic manipulation of passage-cell formation using the mutants and genetic tools developed in this study could be used to investigate the spectrum of nutrients that might be transported through passage cells, as well as to assess the cells' role in nutrient uptake. Another area for future research would be to investigate whether these cells are a potential entry point for microorganisms that cause disease, or those that form a mutually beneficial relationship with plants².

It is interesting that the key genes modulating the cytokinin and auxin fluxes involved in passage-cell formation also act in the patterning of the embryonic root⁸ and water-conducting tissues⁹. Plant development is a continuous process that probably requires developing structures to integrate information from hormonal signalling pathways acting at different times and in different parts of the plant. A key question is how the dynamics of the hormonal signalling interactions that drive so many aspects of plant development are integrated to link these developmental processes. ■

Sedeer el-Showk is a freelance science writer based in Ain Aouda, Morocco.

Ari Pekka Mähönen is at the Institute of Biotechnology, HiLIFE, University of Helsinki, Helsinki 00014, Finland, and in the Faculty of Biological and Environmental Sciences, University of Helsinki.

e-mails: sedeer@elshowk.com; aripekka.mahonen@helsinki.fi

1. Kroemer, H. *Wurzelhaut: Hypodermis und Endodermis der Angiospermenwurzel* (Nägele, 1903).
2. Peterson, C. A. & Enstone, D. E. *Physiol. Plant.* **97**, 592–598 (1996).
3. Wu, H., Jaeger, M., Wang, M., Li, B. & Zhang, B. G. *Ann. Bot.* **107**, 843–853 (2011).
4. Robbins, N. E. II, Trontin, C., Duan, L. & Dinneny, J. R. *Plant Physiol.* **166**, 551–559 (2014).
5. Andersen, T. G. *et al. Nature* **555**, 529–533 (2018).
6. Barberon, M. *et al. Cell* **164**, 447–459 (2016).
7. Mähönen, A. P. *et al. Science* **311**, 94–98 (2006).
8. De Rybel, B. *et al. Science* **345**, 1255215 (2014).
9. Bishopp, A. *et al. Curr. Biol.* **21**, 917–926 (2011).
10. el-Showk, S. *et al. PLoS Comput. Biol.* **10**, e1004450 (2015).
11. Hamburger, D., Rezzonico, E., MacDonald-Comber, Pététot, J., Somerville, C. & Poirier, Y. *Plant Cell* **14**, 889–902 (2002).

This article was published online on 14 March 2018.

CORRECTION

The News & Views 'A surprising chill before the cosmic dawn' by Lincoln Greenhill (*Nature* **555**, 38–39; 2018) incorrectly said in the seventh paragraph that a putative absorption signal looks back as far as 180 million years ago. This should have said "as far back as 180 million years after the Big Bang".

Developmental diversification of cortical inhibitory interneurons

Christian Mayer^{1,2,3,4,*}, Christoph Hafemeister^{2*}, Rachel C. Bandler^{1*}, Robert Machold¹, Renata Batista Brito^{1,5}, Xavier Jaglin^{1,3}, Kathryn Allaway^{1,3}, Andrew Butler^{2,6}, Gord Fishell^{1,3,4,7} & Rahul Satija^{2,6}

Diverse subsets of cortical interneurons have vital roles in higher-order brain functions. To investigate how this diversity is generated, here we used single-cell RNA sequencing to profile the transcriptomes of mouse cells collected along a developmental time course. Heterogeneity within mitotic progenitors in the ganglionic eminences is driven by a highly conserved maturation trajectory, alongside eminence-specific transcription factor expression that seeds the emergence of later diversity. Upon becoming postmitotic, progenitors diverge and differentiate into transcriptionally distinct states, including an interneuron precursor state. By integrating datasets across developmental time points, we identified shared sources of transcriptomic heterogeneity between adult interneurons and their precursors, and uncovered the embryonic emergence of cardinal interneuron subtypes. Our analysis revealed that the transcription factor *Mef2c*, which is linked to various neuropsychiatric and neurodevelopmental disorders, delineates early precursors of parvalbumin-expressing neurons, and is essential for their development. These findings shed new light on the molecular diversification of early inhibitory precursors, and identify gene modules that may influence the specification of human interneuron subtypes.

Cortical inhibitory neurons are a diverse population that varies widely in morphology, connectivity and patterns of activity¹. This group of neurons is developmentally derived from progenitors located in embryonic proliferative zones known as the medial, caudal and lateral ganglionic eminences (MGE, CGE and LGE, respectively)¹. Although each eminence gives rise to non-overlapping types of interneurons, the genetic programs driving interneuron fate specification and maintenance are not well understood. Diversity is first apparent in the regional expression of a limited number of transcription factors within the ganglionic eminences^{2,3}. For example, the transcription factor *Nkx2-1* is expressed throughout the entire MGE, but is not expressed in the CGE or LGE⁴, whereas the transcription factor *Lhx8* is expressed only within a subdomain of the MGE². However, how these early sources of heterogeneity generate the vast diversity of adult interneurons remains unclear, a question that is complicated by the fact that the ganglionic eminences also generate numerous subcortical projection neuron types such as the cholinergic cells of the basal ganglia^{5,6}.

Here we combine multiple single-cell RNA sequencing (scRNA-seq) approaches with genetic fate-mapping techniques to explore the emergence of cellular heterogeneity during early mouse development. Within mitotic progenitors, we found a highly conserved maturation trajectory, accompanied by eminence-specific transcription factor expression that seeds the emergence of later cell diversity. Alongside the exit from the cell cycle, we reconstructed bifurcations into three distinct precursor states, which were highly correlated across eminences, and included a cortical interneuron ground state. Lastly, guided by the genetic diversity seen in mature populations, we connected the transcriptomic heterogeneity of adult interneurons with their embryonic precursors. Our integrated longitudinal analysis reveals the emergence of interneuron subtype identity during development, and identifies genetic regulators responsible for these fate decisions.

Transcriptional profiling of ganglionic eminence cells

We manually dissected ganglionic eminence cells from wild-type mouse embryos at embryonic day (E)13.5 (for the MGE) or E14.5 (for the CGE and LGE)—time points corresponding to peak neurogenesis in these structures^{7,8}, which include both dividing mitotic progenitors as well as postmitotic precursor cells (Fig. 1a, Supplementary Table 1). After cell dissociation, we used droplet-based single-cell mRNA sequencing (Drop-seq)⁹ to sequence the transcriptomes of 5,622 single cells from the MGE, 7,401 from the CGE and 8,543 from the LGE, from replicate experiments, observing on average 1,626 unique molecular identifiers (UMIs) per cell. We performed latent variable regression to mitigate heterogeneity resulting from cell-cycle state^{10,11} (Extended Data Fig. 1)—preventing subsequent analysis from being dominated by mitotic phase-specific gene expression—and filtered out rare contaminating populations of excitatory neurons (*Neurod6*; 2.6% of cells) and endothelial cells (*Igfbp7*; 0.7% of cells) (Fig. 1b, c). The remaining 96.7% of cells were neuronal progenitors and precursors derived from the ganglionic eminences (for example, *Dlx1*; Fig. 1b, c). Within this population, the expression of early, intermediate and late marker genes was strongly associated with the top diffusion map coordinates (DMC; Extended Data Fig. 1). To establish a quantitative temporal account of differentiation programs within each eminence, we fit a principal curve through the DMC, representing an ordered ‘maturation trajectory’ for single cells based on their expression profiles¹² (Fig. 1d). We obtained very similar trajectories using approaches based on principal component analysis or reverse graph embedding¹³ (Extended Data Fig. 1), and observed that the maturation trajectory recapitulated known dynamics associated with neuronal maturation (Fig. 1e) while also segregating ganglionic eminence cells into mitotic and postmitotic phases (Fig. 1f, Extended Data Fig. 1). To independently confirm the association of the maturation trajectory with real time, we used FlashTag technology¹⁴ to fluorescently label cells in the ventricular zone¹⁵ of the

¹NYU Neuroscience Institute, Langone Medical Center, New York, New York 10016, USA. ²New York Genome Center, New York, New York 10013, USA. ³Harvard Medical School, Department of Neurobiology, Boston, Massachusetts 02115, USA. ⁴Broad Institute, Stanley Center for Psychiatric Research, Cambridge, Massachusetts 02142, USA. ⁵Dominick P Purpura Department of Neuroscience, Albert Einstein College of Medicine, Bronx, New York 10461, USA. ⁶Center for Genomics and Systems Biology, New York University, New York, New York 10012, USA. ⁷Center for Genomics and Systems Biology, New York University, PO Box 129188, Saadiyat Island, Abu Dhabi, United Arab Emirates.

*These authors contributed equally to this work.

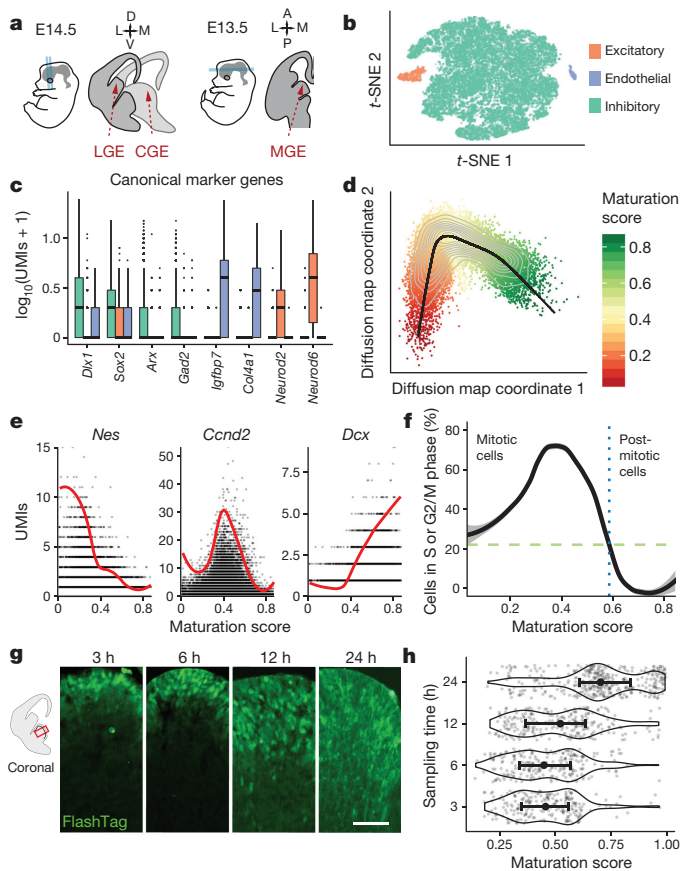


Figure 1 | Transcriptional landscape of single cells in the ganglionic eminences. **a**, Schematic of experimental workflow. Axes: A, anterior; D, dorsal; L, lateral; M, medial; P, posterior; V, ventral. **b**, Visualization of Drop-seq data from ganglionic eminence cells using *t*-distributed stochastic neighbour embedding. **c**, Canonical marker expression in ganglionic eminence precursors, excitatory neurons and vascular endothelial cells; colours as in **b**. **d**, A principal curve was fitted to the dominant diffusion map coordinates to order cells along a maturation trajectory. **e**, Expression (molecules per cell) of canonical regulators, as a function of the position along the maturation trajectory. The curve reflects local averaging of single-cell expression. Locally averaged values were multiplied by five for visualization on the same scale as the molecule counts. **f**, Percentage of cycling cells as a function of the position along the maturation trajectory; the dotted blue line marks the inferred mitotic-to-postmitotic transition. **g**, Coronal brain sections of the ganglionic eminences, as cells migrate away from the ventricular zone (the apical surface of the ventricular zone is at the top of the images). Images were taken 3, 6, 12 and 24 h after fluorescent labelling with FlashTag technology. Scale bar, 50 μ m. **h**, Maturation score distributions of FlashTag labelled cells, separated by time point.

ganglionic eminences, and performed scRNA-seq on cohorts of 3-, 6-, 12- and 24-h-old neurons as they migrated away from the ventricle (Supplementary Table 1, Fig. 1g). As expected, neurons generated at these sequential time points were distributed progressively along the maturation trajectory timeline (Fig. 1h, Extended Data Fig. 1).

The MGE and CGE are known to produce non-overlapping types of cortical interneurons¹⁶. To identify regionally expressed transcription factors^{2,3,17}, we performed a differential expression analysis and found a small number of genes for transcription factors that were enriched in mitotic progenitors within particular eminences (Fig. 2a, Extended Data Fig. 2, Supplementary Table 2), many of which (for example, *Nr2f1*, *Nr2f2*, *Nkx2-1*) have previously been characterized⁶. Next, we identified the sequential patterns of gene expression characterizing the initial stages of cell differentiation. The majority of dynamically expressed genes followed robust and highly reproducible sequential waves of gene expression in all three eminences (Fig. 2b, Extended Data

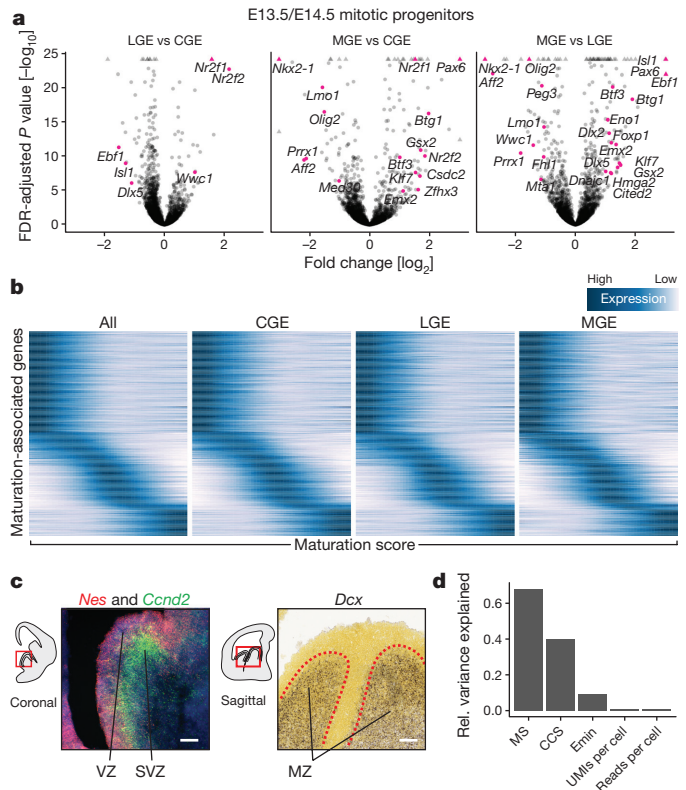


Figure 2 | A common developmental program of gene expression functions in the mitotic progenitors of all three ganglionic eminences. **a**, Volcano plots depicting differential gene expression across eminences for early mitotic cells (maturation score < 0.3). Transcription factors are annotated. **b**, Gene-expression dynamics in mitotic cells, based on local averaging of single-cell data, plotted along maturation score for 840 developmentally regulated genes that were conserved across eminences. **c**, ISH patterns of early, intermediate and late maturation-trajectory genes in the ganglionic eminences that are highly expressed within anatomical boundaries of the ventricular zone, subventricular zone and mantle zone, respectively. **d**, The variance explained individually by a set of annotated factors, relative to the variance explained by the first principal component. Calculated independently for maturation score (MS), cell cycle score (CCS), eminence of origin (Emin), UMIs per cell and reads per cell.

Fig. 2d, Supplementary Table 3). *In situ* hybridization (ISH) confirmed that these waves describe the sequential expression of stem-cell (for example, *Nes*), proneural (for example, *Ascl1*) and neurogenic genes (for example, *Dcx*), approximately correlating with the spatiotemporal progression from the ventricular zone to the mantle zone (Fig. 2c, Extended Data Fig. 3). Developmental progression and cell cycle were the primary sources of transcriptional variance in these progenitors (Supplementary Methods), with maturation proportionally explaining sixfold more variance compared to eminence-of-origin (Fig. 2d).

To detect the potential fate divergence of cells along the maturation trajectory, we bootstrapped the construction of a minimum spanning tree¹⁸ (Fig. 3a, Supplementary Methods), and summarized the combined result using multidimensional scaling. We first observed evidence of clear fate bifurcations as cells become postmitotic, and precursors from all ganglionic eminences branched into distinct precursor states (Fig. 3b, Supplementary Methods). The sequencing of MGE progenitors at substantially higher depth with plate-based scRNA-seq revealed no transcriptomic evidence of similar bifurcations within mitotic cells (Extended Data Fig. 4a–c; Supplementary Table 1). Moreover, when we performed the unsupervised branching analysis only in mitotic progenitors, we found no evidence for the specification of distinct interneuron fates. Instead, consistent with our previous analysis of the maturation trajectory, heterogeneity was driven primarily by maturation state or cell cycle, which may reflect the existence of mitotic

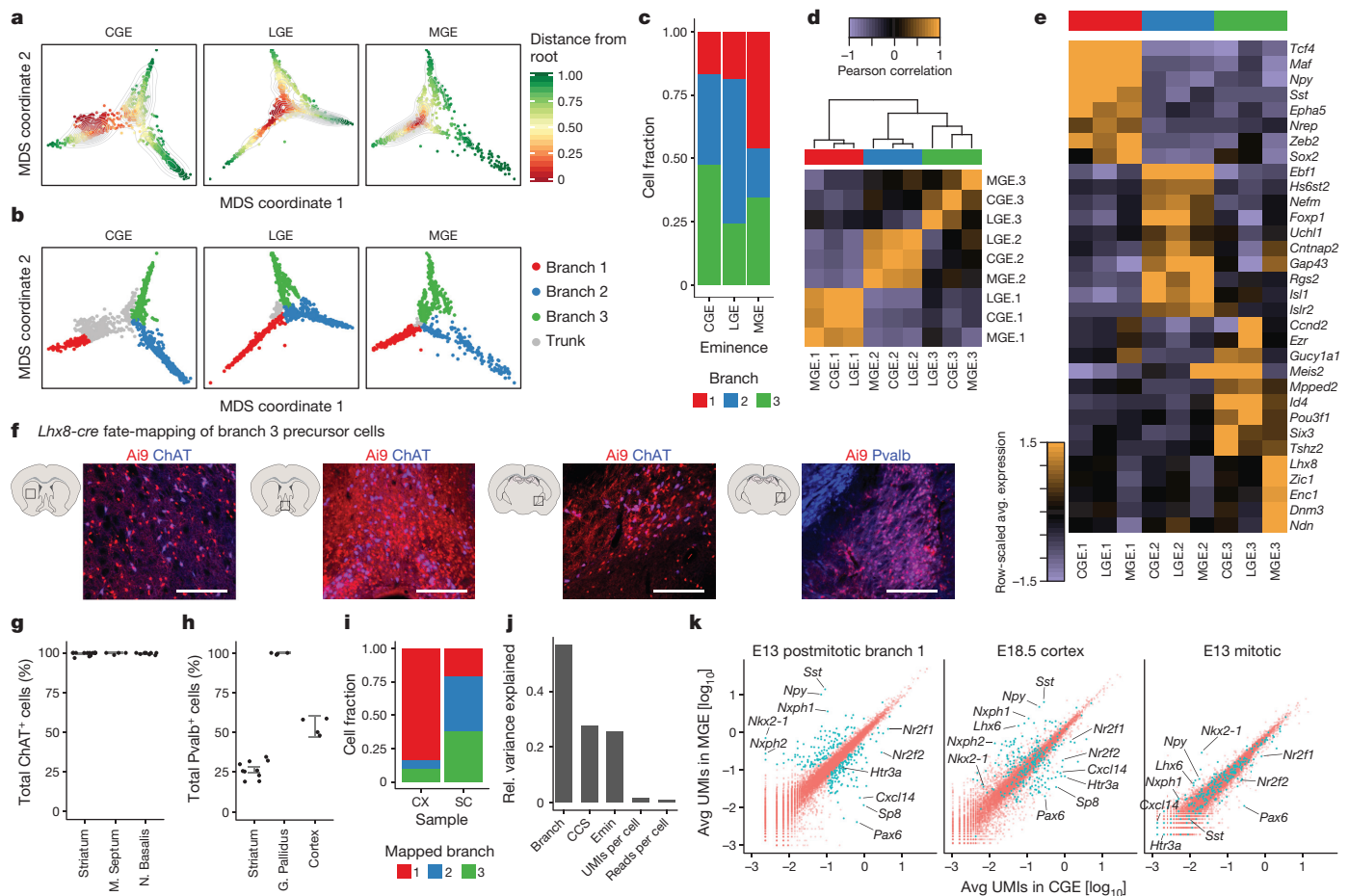


Figure 3 | Postmitotic cells from all eminences pass through distinct precursor states. a, Multidimensional scaling (MDS) based on the average distance along bootstrapped minimum spanning trees. **b**, Minimum spanning tree traversal assigned cells to the trunk and one of three branches. **c**, Quantitative contributions of cells per branch plotted for each ganglionic eminence. **d**, Hierarchical clustering of branch gene expression correlation. Gene expression was averaged for cells from the same ganglionic eminence and branch. **e**, Heat map depicting the top transcriptomic markers for each branch. **f**, Co-localization of Ai9 from *Lhx8-cre*;Ai9 mice with, from left to right, choline acetyltransferase (ChAT) in the striatum, medial septum, and nucleus basalis, and Pvalb in the globus pallidus. Scale bars, 300 μ m. **g**, Percentage of total ChAT⁺ cells labelled with tdTomato in *Lhx8-cre*;Ai9 mice. $n = 15$ brain sections (striatum), $n = 4$ (medial septum), $n = 8$ (nucleus basalis); 2 mice. **h**, The percentage of total Pvalb⁺ cells labelled with tdTomato in *Lhx8-cre*;Ai9

progenitors undergoing direct and indirect neurogenesis within the ventricular and subventricular zones¹⁹ (Extended Data Fig. 4d–f). Nonetheless, we cannot fully exclude the possibility of earlier fate-determination in mitotic progenitors.

We assigned cells to branches by traversing the final minimum spanning tree and annotating major splits (Fig. 3b, c). Notably, even though branched trajectories for each eminence were calculated independently, branch gene expression markers were highly correlated across eminences (Fig. 3d, e). This indicates that, although each ganglionic eminence generates different cell populations, upon becoming postmitotic, cells from all eminences pass through conserved precursor states. One group of highly correlated branches (precursor state 1) expressed known regulators of interneuron development (*Arx*, *Maf*; Fig. 3e, Supplementary Table 4), whereas a second group of branches (precursor state 2) expressed known projection neuron marker genes (*Isl1*, *Ebf1*; Fig. 3e, Supplementary Table 4). The third group of branches (precursor state 3) exhibited weaker correlation across eminences, with the transcription factor *Lhx8* representing a marker gene for the MGE

mice. $n = 10$ brain sections (striatum), $n = 5$ (globus pallidus), $n = 4$ (cortex), 2 mice. Error bars in **g** and **h** indicate standard deviation across all quantified sections. **i**, Mapping of E18.5 cortical (CX) and subcortical (SC) cells to E13.5 (MGE) or E14.5 (CGE, LGE) branches based on marker gene expression correlations. **j**, Relative variance explained individually by annotated factors for postmitotic cells at E13.5 or E14.5 (branch, CCS, Emin, UMIs per cell and reads per cell) relative to the variance explained by the first principle component. Residual cell cycle variation is due to our conservative cutoff for the mitotic–postmitotic transition. **k**, Differential expression analysis between MGE and CGE postmitotic cells in the interneuron precursor state at E13.5 or E14.5 (left). These genes tend to remain differentially expressed between MGE and CGE-derived populations at E18.5 (middle), which is not the case in E13.5 mitotic progenitors (right); differentially expressed genes are depicted in blue.

branch 3 (Fig. 3e). Genetic fate-mapping using *Lhx8-cre* suggested that neurons within this branch account for the majority, if not all, of the cholinergic projection (nucleus basalis, medial septum) and cholinergic interneuron (striatum) populations, as well as the majority of parvalbumin (Pvalb)-positive projection neurons in the globus pallidus^{5,6} (Fig. 3f–h, Extended Data Fig. 4g).

Diversity emerges from a common precursor state

To confirm that cells passing through precursor state 1 give rise to cortical interneurons, we used genetic fate-mapping strategies to enrich for postmitotic cells derived from ganglionic eminences at E18.5 for scRNA-seq (Supplementary Methods, Extended Data Fig. 4). Using a correlation-based distance metric (Supplementary Methods) we found that, as expected²⁰, more than 80% of *Dlx6a-cre* fate-mapped cortical cells at E18.5 were assigned to precursor state 1, on the basis of their expression of canonical regulators of interneuron development (Fig. 3i, Extended Data Fig. 5). The remaining *Dlx6a-cre* fate-mapped cortical population were assigned to precursor states 2 and 3 (Fig. 3i,

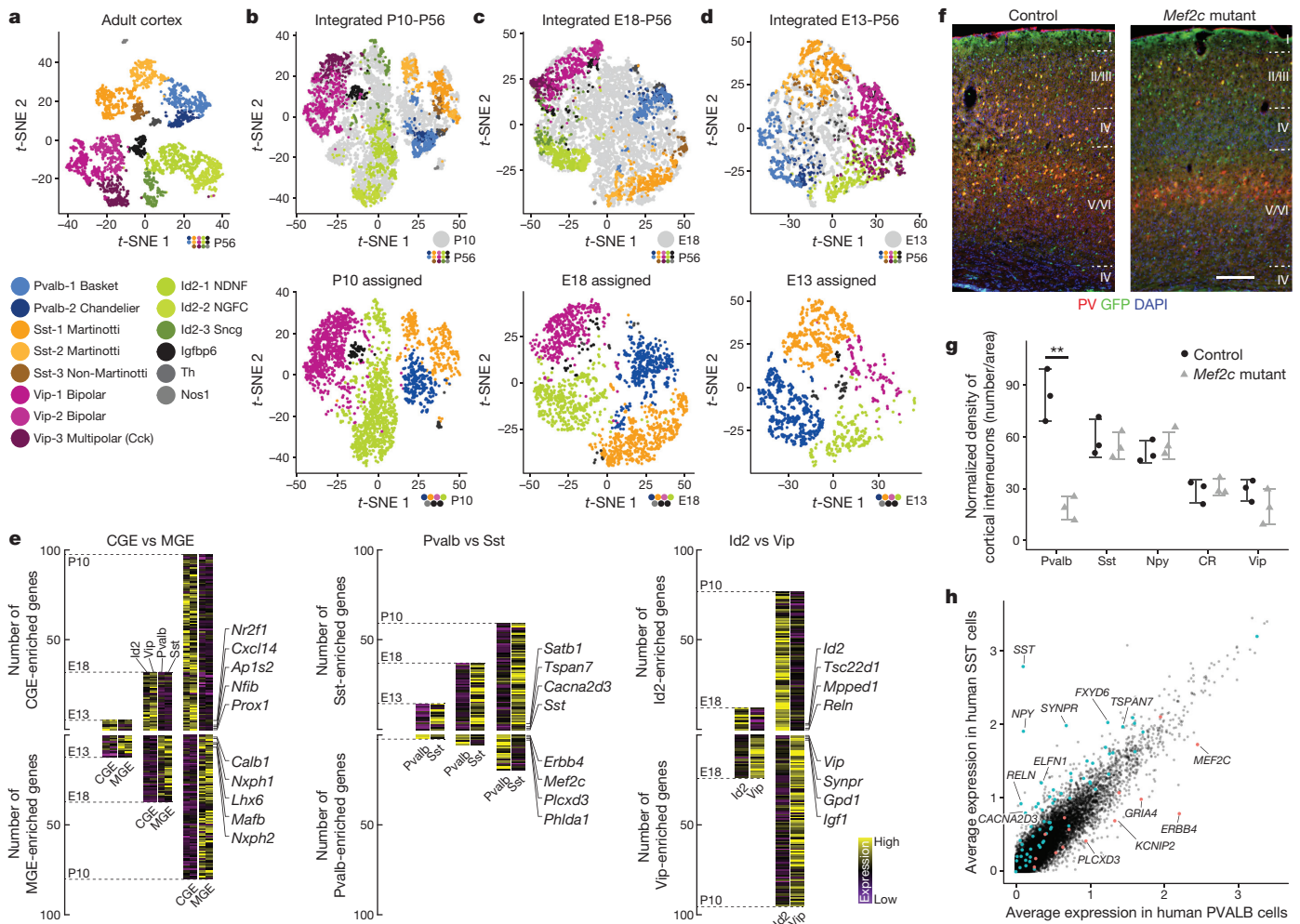


Figure 4 | Integrating developmental scRNA-seq datasets to link embryonic heterogeneity with adult interneuron subtypes. **a**, Graph-based clustering of interneurons from the adult mouse visual cortex (data from ref. 22 and Allen Cell Types Database, <http://celltypes.brain-map.org/download> (2015)). Cluster names denote cardinal classes associated with canonical markers (for example Pvalb or Sst), with additional enumerated subdivisions (for example, Vip-1 or Vip-2). **b–d**, Integration of P10 (**b**), E18.5 (**c**), E13.5 (**d**) precursors with P56 cortical interneurons based on shared sources of variation. Top, adult cells coloured by subtype and precursor cells in grey. Bottom, precursor cells coloured by adult cardinal types to which they are assigned; blue, Pvalb; orange, Sst; violet, Vip; green, Id2; shades of gray, Igfbp6, Th or Nos1. **e**, Differentially expressed genes between CGE- and MGE-derived subsets (left); these genes are conserved in both developmental and adult cells. Each conserved gene is placed on the respective heat map when it is first observed to be differentially expressed during development. The same analysis was performed for Pvalb

Extended Data Fig. 5), probably including the *Meis2*-expressing CGE-derived GABAergic population that has recently been described²¹ (Extended Data Fig. 5). Comparison of the expression profiles of cortical interneuron precursors (precursor state 1) from the MGE and CGE revealed differentially expressed genes, the expression patterns of which are largely maintained in the cortex at later time points (Fig. 3k). Consistently, branching trajectories represented the most important source of variation in these cells, with an increasing contribution attributable to eminence of origin compared to mitotic progenitors (Fig. 3j). Thus, our data reveal how postmitotic pan-eminence transcriptional programs (precursor states) emerge, and in parallel, eminence-specific transcriptional programs escalate.

We next asked when subtype-specific gene expression patterns first appear during interneuron development. In the adult mouse, using a

and Sst subsets (middle), and Vip and Id2 subsets (right). **f**, Conditional deletion of *Mef2c* in inhibitory neurons using *Dlx6a-cre;Mef2c^{loxP/loxP}RCE*. Immunostaining of P20–P22 somatosensory cortex using anti-GFP (green) and anti-Pvalb (red); counterstaining with 4,6-diamidino-2-phenylindole (DAPI) shows cortical layers. Scale bar, 200 μ m. **g**, Density quantification of cortical interneuron subtypes in the P21 somatosensory cortex using antibodies for Pvalb, Sst, Vip, Npy, and calretinin (CR). Error bars reflect s.e.m.; two-tailed unpaired *t*-test, $^{**}P < 0.01$; $n = 3$ brains each for cortical knockout and control. Error bars reflect s.e.m.; two-tailed unpaired *t*-test, $^{**}P < 0.01$; $n = 3$ brains each for cortical knockout and control. **h**, Scatter plot comparing average expression of GABAergic single nuclei from post-mortem human neurons after segregation into Pvalb and Sst types. Each dot represents the expression of a human gene. Markers of embryonic cardinal types are shown in green or blue dots, with a subset of gene names annotated.

publicly available dataset²² (© 2015 Allen Institute for Brain Science, Allen Cell Types Database, available from: <http://celltypes.brain-map.org/download>) we identified 14 inhibitory interneuron subpopulations that encompass known anatomically and physiologically defined subtypes^{23,24} (Fig. 4a, Supplementary Methods, Extended Data Fig. 6). These could be allocated into non-overlapping cardinal types of cortical interneurons (Pvalb, Sst, Vip, Id2, Th, Nos1, Igfbp6). We reasoned that if we could identify heterogeneous gene modules in developing cells that were shared with adult interneurons, we could identify early patterns of specification in precursors. We therefore applied our recently developed tool for the pairwise integration of scRNA-seq datasets^{25,26} (Fig. 4b–d), which ‘aligns’ cell types across datasets based on conserved sources of variation as identified by canonical correlation analysis. This procedure therefore links the heterogeneity observed in

adult cells with heterogeneity in their precursors. On the basis of this alignment, P10 cells exhibited strong evidence of transcriptomic separation beyond cardinal types (Fig. 4b), including clear segregation between Sst Martinotti and non-Martinotti (X94), Vip bipolar and multipolar, and Id2 neurogliaform and non-neurogliaform interneuron subtypes (Fig. 4e, Extended Data Figs 7–9).

Embryonic stages also displayed strong evidence of interneuron specification. On examination of the earliest stages, we observed a separation of Pvalb- and Sst-precursor cells within the E13.5 post-mitotic populations (Fig. 4d), and identified transcriptomic markers that were conserved into adulthood (early marker genes for Pvalb neurons: *Mef2c*, *ErbB4*, *Plcx3*; early marker genes for Sst neurons: *Sst*, *Tspan7*, *Satb1*; Fig. 4e, Extended Data Fig. 7). A minority of E13.5 cells also mapped to Vip and Id2 subsets, but conserved transcriptomic markers did not reach statistical significance until E18.5 (E18.5 markers of Vip neurons: *Vip*, *Synpr*, *Igf1*; E18.5 markers of Id2 neurons: *Reln*, *Mpped1*, *Id2*). By E18.5, all cardinal types of interneurons could be identified, and additional subtypes appeared to be transcriptionally specified as well (Fig. 4e, Extended Data Fig. 9). Notably, segregation into subtypes became evident at different developmental stages. For example, the clear emergence of Sst, Vip and Id2 subtypes was apparent for a subset of cells at E18.5 (Extended Data Fig. 8), but we were unable to clearly subdivide Pvalb neurons by P10, in accordance with their late maturation²⁷. The results of our integrated analyses were in agreement with independent unsupervised analysis of each developmental stage (Extended Data Fig. 8). Consistent with our earlier findings (Fig. 2), we did not observe common sources of variation shared between adult interneurons and mitotic progenitors.

In addition to observing the potential specification of embryonic precursors, our list of cardinal type and subtype markers that are conserved from the ganglionic eminences through adulthood suggests a set of genetic regulators that may have important roles in this process. For example, the gene encoding the transcription factor *Mef2c* was among those that discriminated early Pvalb-precursors from other MGE-derived interneuron types (Fig. 4e). Genome-wide association studies have linked mutations in this gene to Alzheimer's disease, schizophrenia and other neurodevelopmental disorders²⁸. Consistent with our predictions, conditional deletion of *Mef2c* in inhibitory neurons led to a specific loss of Pvalb-interneurons by P20 in cortical layers 2–6 (Fig. 4f, g, Extended Data Fig. 10), indicating that *Mef2c* is essential for the generation of this population. Notably, when examining a published single-nucleus RNA-seq dataset of human post mortem tissue²⁹, we found that a subset of embryonic cardinal type markers from our mouse dataset (including *Mef2c*) was also differentially expressed in adult human interneurons (Fig. 4h). Therefore, the genes we identified as defining embryonic cardinal types are candidates for the regulation of interneuron fate determination and maintenance across species.

Discussion

Our work reveals how subtype-specific heterogeneity progresses from the expression of cardinal genes in progenitors to the emergence of specific subtypes that populate the mature cortex. Postmitotic cells in the ganglionic eminences branch into distinct precursor states, representing populations fated to give rise to interneurons or projection neurons. It seems probable that the superimposition of precursor-state genes and eminence-specific genes act coordinately to bestow the common and unique characteristics within particular GABAergic populations, respectively.

Consequently, precursor genes are likely to direct the developmental cascade and acquisition of general properties that are shared within a given type. This probably ensures, for instance, that interneurons migrate tangentially to the cortex or the hippocampus, whereas projection neurons remain positioned ventrally and form long-range projections. Supplementing these more general programs are the eminence-specific genes that, for example, may direct the axons of

parvalbumin cortical interneurons to form perisomal baskets and the efferents of somatostatin cortical interneurons to reliably target dendrites. These distinct differentiation modules reflect the major cardinal types of cortical interneuron precursors.

The identification of early precursors offers insight into how specific cell types emerge and provides genetic access to immature cortical interneuron subtypes. To broaden the implications of these results, our findings indicate that components of the transcriptional networks underlying interneuron fate specification are conserved between mouse and human, including *Mef2c* and other genes associated with neuropsychiatric disorders. This highlights the power of combining single-cell genomics with analytical tools to identify genes that have important functional roles in the establishment and maintenance of interneuron fates. Our findings mark an initial but important step towards the goal of ultimately linking specific genes to their aetiology in neurodevelopmental and neuropsychiatric disorders.

Online Content Methods, along with any additional Extended Data display items and Source Data, are available in the online version of the paper; references unique to these sections appear only in the online paper.

Received 30 January 2017; accepted 12 February 2018.

Published online 5 March 2018.

- Kepecs, A. & Fishell, G. Interneuron cell types are fit to function. *Nature* **505**, 318–326 (2014).
- Flames, N. *et al.* Delineation of multiple subpallial progenitor domains by the combinatorial expression of transcriptional codes. *J. Neurosci.* **27**, 9682–9695 (2007).
- Yun, K., Garel, S., Fischman, S. & Rubenstein, J. L. R. Patterning of the lateral ganglionic eminence by the *Gsh1* and *Gsh2* homeobox genes regulates striatal and olfactory bulb histogenesis and the growth of axons through the basal ganglia. *J. Comp. Neurol.* **461**, 151–165 (2003).
- Shimamura, K., Hartigan, D. J., Martinez, S., Puelles, L. & Rubenstein, J. L. Longitudinal organization of the anterior neural plate and neural tube. *Development* **121**, 3923–3933 (1995).
- Nóbrega-Pereira, S. *et al.* Origin and molecular specification of globus pallidus neurons. *J. Neurosci.* **30**, 2824–2834 (2010).
- Zhao, Y. *et al.* The LIM-homeobox gene *Lhx8* is required for the development of many cholinergic neurons in the mouse forebrain. *Proc. Natl Acad. Sci. USA* **100**, 9005–9010 (2003).
- Inan, M., Welagen, J. & Anderson, S. A. Spatial and temporal bias in the mitotic origins of somatostatin- and parvalbumin-expressing interneuron subgroups and the chandelier subtype in the medial ganglionic eminence. *Cereb. Cortex* **22**, 820–827 (2012).
- Miyoshi, G. *et al.* Genetic fate mapping reveals that the caudal ganglionic eminence produces a large and diverse population of superficial cortical interneurons. *J. Neurosci.* **30**, 1582–1594 (2010).
- Macosko, E. Z. *et al.* Highly parallel genome-wide expression profiling of individual cells using nanoliter droplets. *Cell* **161**, 1202–1214 (2015).
- Buettner, F. *et al.* Computational analysis of cell-to-cell heterogeneity in single-cell RNA-sequencing data reveals hidden subpopulations of cells. *Nat. Biotechnol.* **33**, 155–160 (2015).
- Tirosh, I. *et al.* Dissecting the multicellular ecosystem of metastatic melanoma by single-cell RNA-seq. *Science* **352**, 189–196 (2016).
- Petropoulos, S. *et al.* Single-cell RNA-seq reveals lineage and X chromosome dynamics in human preimplantation embryos. *Cell* **165**, 1012–1026 (2016); erratum 165, 1012–1026, (2016).
- Qiu, X. *et al.* Reversed graph embedding resolves complex single-cell trajectories. *Nat. Methods* **14**, 979–982 (2017).
- Quah, B. J. C. & Parish, C. R. The use of carboxyfluorescein diacetate succinimidyl ester (CFSE) to monitor lymphocyte proliferation. *J. Vis. Exp.* <https://doi.org/10.3791/2259> (2010).
- Telley, L. *et al.* Sequential transcriptional waves direct the differentiation of newborn neurons in the mouse neocortex. *Science* **351**, 1443–1446 (2016).
- Rudy, B., Fishell, G., Lee, S. & Hjerling-Leffler, J. Three groups of interneurons account for nearly 100% of neocortical GABAergic neurons. *Dev. Neurobiol.* **71**, 45–61 (2011).
- Waclaw, R. R., Ehrman, L. A., Pierani, A. & Campbell, K. Developmental origin of the neuronal subtypes that comprise the amygdalar fear circuit in the mouse. *J. Neurosci.* **30**, 6944–6953 (2010).
- Trapnell, C. *et al.* The dynamics and regulators of cell fate decisions are revealed by pseudotemporal ordering of single cells. *Nat. Biotechnol.* **32**, 381–386 (2014).
- Petros, T. J., Bultje, R. S., Ross, M. E., Fishell, G. & Anderson, S. A. Apical versus basal neurogenesis directs cortical interneuron subclass fate. *Cell Rep.* **13**, 1090–1095 (2015).
- Wichterle, H., Turnbull, D. H., Nery, S., Fishell, G. & Alvarez-Buylla, A. *In utero* fate mapping reveals distinct migratory pathways and fates of neurons born in the mammalian basal forebrain. *Development* **128**, 3759–3771 (2001).

21. Frazer, S. *et al.* Transcriptomic and anatomic parcellation of 5-HT_{3A}R expressing cortical interneuron subtypes revealed by single-cell RNA sequencing. *Nat. Commun.* **8**, 14219 (2017).
22. Tasic, B. *et al.* Adult mouse cortical cell taxonomy revealed by single cell transcriptomics. *Nat. Neurosci.* **19**, 335–346 (2016).
23. Tremblay, R., Lee, S. & Rudy, B. GABAergic interneurons in the neocortex: from cellular properties to circuits. *Neuron* **91**, 260–292 (2016).
24. Petilla Interneuron Nomenclature Group. Petilla terminology: nomenclature of features of GABAergic interneurons of the cerebral cortex. *Nat. Rev. Neurosci.* **9**, 557–568 (2008).
25. Butler, A. & Satija, R. Integrated analysis of single cell transcriptomic data across conditions, technologies, and species. Preprint at <https://www.biorxiv.org/content/early/2017/07/18/164889> (2017).
26. Butler, A., Hoffman, P., Smibert, P., Papalexi, E. & Satija, R. Integrating single-cell transcriptomic data across different conditions, technologies, and species. *Nat. Biotechnol.* <https://doi.org/10.1038/nbt.4096> (2018).
27. Alcántara, S., de Lecea, L., Del Río, J. A., Ferrer, I. & Soriano, E. Transient colocalization of parvalbumin and calbindin D28k in the postnatal cerebral cortex: evidence for a phenotypic shift in developing nonpyramidal neurons. *Eur. J. Neurosci.* **8**, 1329–1339 (1996).
28. Harrington, A. J. *et al.* MEF2C regulates cortical inhibitory and excitatory synapses and behaviors relevant to neurodevelopmental disorders. *eLife* **5**, 140 (2016).
29. Habib, N. *et al.* Massively parallel single-nucleus RNA-seq with DroNc-seq. *Nat. Methods* **14**, 955–958 (2017).

Supplementary Information is available in the online version of the paper.

Acknowledgements We thank members of the Fishell and Satija laboratories, and C. Desplan, for feedback and discussion; L. Harshman, B. Bracken and W. Stephenson for assistance with scRNA-seq experiments; and N. Habib for assistance with published datasets. This work was supported by National Institutes of Health (NIH) grants R01 NS074972 (G.F.), R01 NS081297 (G.F.), MH071679-12 (G.F. and R.S.), NIH DP2-HG-009623 (R.S.), European Molecular Biology Organization ALTF 1295-2012 (C.M.), Deutsche Forschungsgemeinschaft Postdoctoral Fellow (C.H.), NIH F30MH114462 (R.C.B.), T32GM007308 (R.C.B.), NIH F31NS103398 (K.A.), and National Science Foundation DGE1342536 (A.B.). G.F. is also supported by a grant from the Simons Foundation (274578).

Author Contributions C.M., C.H., R.C.B., G.F. and R.S. conceived the research. C.M. and R.C.B. led experimental work, assisted by R.M., R.B.B., X.J., K.A. and supervised by G.F. C.H. led computational analysis, assisted by C.M. and A.B., and supervised by R.S. All authors participated in interpretation and writing the manuscript.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing interests. Readers are welcome to comment on the online version of the paper. Publisher's note: Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations. Correspondence and requests for materials should be addressed to G.F. (gordon_fishell@hms.harvard.edu) or R.S. (rsatija@nygenome.org).

Reviewer Information *Nature* thanks A. Klein and the other anonymous reviewer(s) for their contribution to the peer review of this work.

METHODS

Data reporting. No statistical methods were used to predetermine sample size. The experiments were not randomized and the investigators were not blinded to allocation during experiments and outcome assessment.

Animals. All mouse colonies were maintained in accordance with protocols approved by the Institutional Animal Care and Use Committee at the NYU School of Medicine. Mouse strains used are the following: wild-type Swiss Webster females (Taconic Biosciences), *Dlx6a*(Tg)-*cre* (ref. 30), *Lhx6*(BAC)-GFP (GENSAT), *Lhx8-cre/cerulean* (Jax stock #023453), *Rosa26^{LSL-tdTomato}* (Ai9) (ref. 31), *Rosa26* (CAG)-*LSL-eGFP* (*RCE^{loxP}*) (Jax stock #32037) (ref. 32), *Dlx5/6*(Tg)-*cre/eGFP* (Jax stock #023724; for characterization information, see images at the Allen Institute for Brain Science website (http://connectivity.brain-map.org/transgenic/imageseries/list/1.html?gene_term=Dlx5-CreERT2)), and *Mef2c^{fl/fl}* (ref. 33). Both male and female mice were used for all single-cell RNA sequencing experiments.

Wild-type Drop-seq experiments. Mouse embryos at 13.5 (MGE) and 14.5 (CGE and LGE) days' gestation were isolated from 6–8-week-old wild-type Swiss Webster timed-pregnant dams ordered from Taconic Biosciences. Embryos were staged in days post coitus, with embryonic day (E) 0.5 defined as noon of the day a vaginal plug was detected after overnight mating. The method of euthanasia for pregnant dams was inhaled-isoflurane overdose, and death was confirmed with decapitation. Surgical access to the uterine horns enabled removal of embryos. After removal from the mother, embryos were stored on ice in Leibovitz's L-15 medium and 1% fetal bovine serum. Brains were removed from the embryos and embedded in 1% ultrapure low melting point agarose and sectioned in 50- μ m sections with a vibratome (Leica VT1200S). The MGE, CGE or LGE were dissected from each embryo. MGEs were dissected from horizontal brain sections, whereas CGEs and LGEs were dissected from coronal brain sections. Tissue from several embryos was pooled together before dissociation.

Single-cell dissociation. Embryonic brain tissue pooled from several embryos was dissociated into a single-cell suspension using a papain dissociation system (Worthington Biochemical) according to the manufacturer's instructions. Postnatal brain tissue was dissociated with 1 mg ml⁻¹ of pronase (Roche, #10 165 921 001) in ice-cold prebubbled artificial cerebrospinal fluid for 25 min.

Fluorescence-activated cell sorting. Fluorescent cells from *Dlx6a-cre;Ai9*, *Dlx6a-cre;RCE^{loxP}*, *Lhx6*(BAC)-GFP and CellTrace injected (FlashTag) brain tissue were sorted on a Sony SY3200 sorter with a 100- μ m nozzle. Cells were sorted in bulk for experiments using Drop-seq and the 10x Genomics platform, whereas for experiments using plate based scRNA-seq methods, single cells were sorted into 96-well plates and immediately frozen on dry ice.

Single-cell RNA sequencing and library preparation. Drop-seq was run on single cells according to the Online Dropseq Protocol v3.1 (December 2015) and the methods published in ref. 9. Drop-seq flow rates (oil: 6,000 μ l h⁻¹, cells: 2,000 μ l h⁻¹, beads: 2,000 μ l h⁻¹) were optimized based on human-mouse species mixing experiments with a 1–2% doublet rate. Libraries were prepared with the Nextera XT DNA Library Preparation Kit according to the manufacturer's instructions. For experiments using the 10x Genomics platform, the Chromium Single Cell 3' Library & Gel Bead Kit v2 (PN-120237), Chromium Single Cell 3' Chip kit v2 (PN-120236) and Chromium i7 Multiplex Kit (PN-120262) were used according to the manufacturer's instructions in the Chromium Single Cell 3' Reagents Kits V2 User Guide.

For single cells sorted into 96-well plates, cells were immediately lysed and mRNAs were released when single cells were sorted into wells with 5 \times Maxima reverse transcription buffer, dNTP mixture, RNase inhibitors (SUPERase RNase Inhibitor, Thermo Fisher Scientific #AM2696) and water. We reverse-transcribed the mRNAs using Superscript II Reverse Transcriptase (Thermo Fisher Scientific #18064071), and amplified cDNAs for each cell in individual wells using the Smart-seq2 (ref. 34) protocol, with a custom modification in which a 12-base cell barcode was included in the 3'-end reverse transcriptase primer. This enabled us to perform multiplexed pooling before library preparation with the Nextera XT DNA sample prep kit (Illumina), and returned 3' biased data similar to the Drop-seq protocol. We quantified the cDNA libraries on an Agilent BioAnalyzer and sequenced them on a HiSeq 2500.

FlashTag. Immediately before use, 10 mM CFSE (Life Technologies, #C34554) CellTrace solution was prepared according to the manufacturer's instructions, and 2–3 μ l was injected into the lateral ventricle of E12.5 or E13.5 wild-type mouse embryos. Embryos were collected 3, 6, 12 and 24 h post-injection.

Fluorescent in situ hybridization. Fluorescent *in situ* hybridization (ISH) for nestin (*Nes*) and cyclin D2 (*Ccnd2*) transcripts (Fig. 2d) was performed as previously described³⁵. Antisense cRNA probes were prepared by T7 polymerase *in vitro* transcription of PCR product templates generated using the following primers: Nestin, 5'-AGCAGTGCCTGGAAGTGAAG-3' and 5'-GCACATTAATACGACTCACTATAGGGCTGGATCCCTCAGCTTGG-3'; Cyclin D2, 5'-ACCTCCCGCAGTGTCTCTA-3' and 5'-AATTAATACGACTCACTATAGGCTGCTCTTACGGAACTGCT-3'

Immunohistochemistry. *Lhx8 fate-mapping.* To fate-map *Lhx8*-lineage neurons, *Lhx8-cre/cerulean* mice were crossed with *Ai9* mice. Offspring with both alleles were then transcardially perfused with PBS followed by 4% paraformaldehyde (PFA) in PBS at postnatal day 21. The brains were collected, fixed overnight in 4% PFA at 4 °C, embedded in 4% agarose, and sectioned at 50 μ m on a Leica VT1200S vibratome before proceeding with immunohistochemistry. *Lhx8-cre/cerulean;Ai9* brain sections were blocked for 1 h in 10% normal donkey serum, 0.3% Triton-X. Sections were incubated at 4 °C overnight in the following primary antibodies: goat anti-ChAT (1:250, Millipore AB144P), goat anti-Pvalb (1:1000, Swant PVG213), rabbit anti-SST (1:3000, Peninsula Labs T-4102), rabbit anti-GABA (1:2000, Sigma A2052), chicken anti-GFP (Aves Labs 1020), rabbit anti-DsRed (Clontech 632496). Following several washes in PBS, sections were incubated at room temperature for 1 h in the following secondary antibodies: Alexa Fluor 488-donkey anti-chicken, Alexa Fluor 594-donkey anti-rabbit, and Alexa Fluor 647-donkey anti-goat and rabbit (all Jackson ImmunoResearch). Sections were then washed again several times in PBS, treated with DAPI as a counterstain, and mounted on slides. Sections were imaged using a Zeiss Axioimager A2 and processed in ImageJ.

Mef2c conditional inactivation. We conditionally inactivated *Mef2c* by crossing the *Mef2c^{fl/fl}* allele with either a *Dlx6a-cre* or *Dlx5/6-cre* driver line. As the recombination mediated by these two driver lines is indistinguishable, they are used interchangeably in Fig. 4f, g, respectively. This results in the deletion of *Mef2c* in all cortical interneurons during embryogenesis, shortly after they become postmitotic. We took advantage of the *RCE^{loxP}* reporter line, which upon Cre-mediated recombination enables the expression of GFP.

Density of cortical-interneuron subtypes in *Mef2c* conditional mutants. Brains from *Dlx5/6-cre;Mef2c^{loxP/loxP}RCE* mice were fixed by transcardial perfusion with 4% PFA in PBS followed by a 1-h post-fixation period on ice with 4% PFA/PBS solution. Brains were rinsed with PBS and cryoprotected using 15% sucrose/PBS solution for 6 h and 30% sucrose/PBS solution overnight at 4 °C. Tissues were embedded in Tissue-Tek, frozen on dry ice, and cryosectioned at thicknesses of 20 μ m. Sections for immunohistochemistry analysis were processed using 1.5% normal goat serum and 0.1% Triton X-100 in all procedures except for washing steps, in which only PBS was used. Sections were blocked for 1 h, followed by incubation with the primary antibodies overnight at 4 °C. Cryostat tissue sections were stained with the primary antibodies rat anti-SST (1:250, Chemicon), mouse anti-Pvalb (1:1,000, Sigma), and rabbit anti-VIP (1:250, ImmunoStar), rabbit anti-Neuropeptide Y (1:500; Incstar), mouse anti-Calretinin (1:1,500; Chemicon). Secondary antibodies conjugated with Alexa fluorescent dyes were applied for 1 h at room temperature to visualize the signals. Nuclear counterstaining was performed with DAPI solution. All analysis was evaluated in the somatosensory cortex. Density quantification for cortical-interneuron subtypes was calculated as number of (peptide marker) + expressing cortical interneurons/area in the P21 control and *Mef2c* cortical knockout somatosensory cortex. To minimize counting bias, we compared sections of equivalent bregma positions, defined according to the Mouse Brain Atlas³⁶. The total number of cells expressing the marker were counted for a defined and normalized optical area. Three brains each were used for knockout and control.

Layer distribution of Pvalb-expressing cortical interneurons in *Mef2c* conditional mutants. Tissue from conditional *Mef2c* mutants (*Dlx6a-cre;Mef2c^{fl/fl}*) and controls (*Dlx6a-cre;Mef2c^{fl/+}*) were analysed at P20–P21. Adult mice were transcardially perfused with 4% PFA after being anaesthetized by intraperitoneal administration of Sleepaway. Brains that were processed for immunofluorescence on slides were post fixed in 4% PFA in PBS at 4 °C and cryopreserved following the perfusion and brain collection. 16- μ m coronal sections were obtained using Cryostat (Leica Biosystems) and collected on super-frost coated slides, then allowed to dry and stored at –20 °C until use. For immunofluorescence, cryosections were thawed and allowed to dry for 5–10 min and rinsed twice in 1 \times PBS. They were incubated at room temperature in a blocking solution of PBST (PBS, 0.1% Triton X-100) and 10% normal donkey serum (NDS) for 60 min, followed by incubation with primary antibodies in PBST and 1% NDS at 4 °C overnight. Primary antibodies are as follows: anti-GFP chicken polyclonal IgY (1:1,000) (Abcam Ab13970), anti-parvalbumin (Pvalb) goat (Swant PVG 213) and anti-parvalbumin (Pvalb) rabbit (Swant PV25). Samples were then washed three times with PBST and incubated with fluorescence conjugated secondary Alexa antibodies (Life Technologies) in PBST with 1% NDS at room temperature for 60–90 min. Slides were then incubated for 30 s with DAPI, washed three times with PBST and once with PBS. Finally, slides were mounted with Fluoromount G (Southern Biotech) and imaged. To quantify the layer distribution and density of various populations of cortical interneurons, the proportion of interneurons of given subtypes over the total number of fate-mapped interneurons across cortical layers was manually determined in ImageJ. Percentages presented in Fig. 4 were calculated by dividing the number of marker-positive neurons in each layer (for example, layer I, layer II/III, layer IV and layer V/VI) by the total number of reporter-positive neurons. Percentages were compared with

repeated *t*-tests in GraphPad Prism, and means \pm s.d. are represented. Three brains each were used for knockout and control, 3–4 sections per brain. See Extended Data Fig. 10a.

Single-cell RNA-seq data processing. The raw Drop-seq data was processed with the standard pipeline (Drop-seq tools version 1.12 from McCarroll laboratory). Reads were aligned to the ENSEMBL release 84 *Mus musculus* genome. 10x Genomics data was processed using the same pipeline as for Drop-seq data, adjusting the barcode locations accordingly.

Data filtration. We selected cells for downstream processing in each Drop-seq run, using the quality control metrics output by the Drop-seq tools package⁹, as well as metrics derived from the UMI matrix. 1) We first removed cells with a low number (<700) of unique detected genes. From the remaining cells, we filtered additional outliers. 2) We removed cells for which the overall alignment rate was less than the mean minus three standard deviations. 3) We removed cells for which the total number of reads (after \log_{10} transformation) was not within three standard deviations of the mean. 4) We removed cells for which the total number of unique molecules (UMIs, after \log_{10} transformation) was not within three standard deviations of the mean. 5) We removed cells for which the transcriptomic alignment rate (defined by PCT_USABLE_BASES) was not within three standard deviations of the mean. 6) We removed cells that showed an unusually high or low number of UMIs given their number of reads by fitting a loess curve (span = 0.5, degree = 2) to the number of UMIs with number of reads as predictor (both after \log_{10} transformation). Cells with a residual more than three standard deviations away from the mean were removed. 7) With the same criteria, we removed cells that showed an unusually high or low number of genes given their number of UMIs. Of these filter steps, step 1 removed the majority of cells. Steps 2 to 7 removed only a small number of additional cells from each eminence (2% to 4%), and these cells did not exhibit unique or biologically informative patterns of gene expression.

Data normalization. The raw data per Drop-seq run is a UMI count matrix with genes as rows and cells as columns. The values represent the number of UMIs that were detected. The aim of normalization is to make these numbers comparable between cells by removing the effect of sequencing depth and biological sources of heterogeneity that may confound the signal of interest, in our case cell cycle stage.

A common approach to correct for sequencing depth is to create a new normalized expression matrix x with $x_{ij} = \log\left(\frac{c_{ij} \times 10,000}{m_j}\right)$, in which c_{ij} is the molecule count of gene i in cell j and m_j is the sum of all molecule counts for cell j . This approach assumes that c_{ij} increases linearly with m_j , which is true only when the set of genes detected in each cell is roughly the same. However, for Drop-seq, in which the number of UMIs is low per cell compared to the number of genes present, the set of genes detected per cell can be quite different. Hence, we normalize the expression of each gene separately by modelling the UMI counts as coming from a generalized linear model with negative binomial distribution, the mean of which can be dependent on technical factors related to sequencing depth. Specifically, for every gene we model the expected value of UMI counts as a function of the total number of reads assigned to that cell, and the number of UMIs per detected gene (sum of UMI divided by number of unique detected genes). To solve the regression problem, we use a generalized linear model (glm function of base R package) with a regularized overdispersion parameter theta. Regularizing theta helps us to avoid overfitting which could occur for genes whose variability is mostly driven by biological processes rather than sampling noise and dropout events. To learn a regularized theta for every gene, we perform the following procedure.

1) For every gene, obtain an empirical theta using the maximum likelihood model (theta.ml function of the MASS R package) and the estimated mean vector that is obtained by a generalized linear model with Poisson error distribution.

2) Fit a line (loess, span = 0.33, degree = 2) through the variance–mean UMI count relationship (both \log_{10} transformed) and predict regularized theta using the fit. The relationship between variance and theta and mean is given by $\text{variance} = \text{mean} + (\text{mean}^2/\text{theta})$.

Normalized expression is then defined as the Pearson residual of the regression model, which can be interpreted as the number of standard deviations by which an observed UMI count was higher or lower than its expected value. Unless stated otherwise, we clip expression to the range [−30, 30] to prevent outliers from dominating downstream analyses.

Removal of cell cycle effect. The normalization method described above aims to reduce the effect of technical factors in scRNA-seq data (primarily, depth) from downstream analyses. However, heterogeneity in cell cycle stage, particularly among mitotic cells transitioning between S and G2/M phases, also can drive substantial transcriptomic variation that can mask biological signal. To mitigate this effect, we use a two-step approach: 1) quantify cell cycle stage for each cell using supervised analyses with known stage-specific markers, 2) regress the effect of cell cycle stage using the same negative binomial regression as outlined above. For the first step we use a previously published list of cell cycle dependent genes (43 S phase genes, 54

G2/M phase genes) for an enrichment analysis similar to that proposed in ref. 11. For each cell, we compare the sum of phase-specific gene expression (\log_{10} transformed UMIs) to the distribution of 100 random background genes sets, where the number of background genes is identical to the phase gene set, and the background genes are drawn from the same expression bins. Expression bins are defined by 50 non-overlapping windows of the same range based on $\log_{10}(\text{mean UMI})$. The phase-specific enrichment score is the expression *z*-score relative to the mean and standard deviation of the background gene sets. Our final ‘cell cycle score’ (Extended Data Fig. 1) is the difference between S-phase score and G2/M-phase score.

For a final normalized dataset with cell cycle effect removed, we perform negative binomial regression with technical factors and cell cycle score as predictors. Although the cell cycle activity was regressed out of the data for downstream analysis, we stored the computed cell cycle score before regression, enabling us to remember the mitotic phase of each individual cell. Notably, our regression strategy is tailored to mitigate the effect of transcriptional heterogeneity within mitotic cells in different phases, and should not affect global differences between mitotic and non-mitotic cells that may be biologically relevant.

Dimensionality reduction. Throughout the manuscript we use diffusion maps, a non-linear dimensionality reduction technique³⁷. We calculate a cell-to-cell distance matrix using $1 - \text{Pearson correlation}$ and use the diffuse function of the diffusionMap R package with default parameters to obtain the first 50 DMCs. To determine the significant DMCs, we look at the reduction of eigenvalues associated with DMCs. We determine all dimensions with an eigenvalue of at least 4% relative to the sum of the first 50 eigenvalues as significant, and scale all dimensions to have mean 0 and standard deviation of 1.

Initial clustering of all cells. To identify contaminating cell populations and assess overall heterogeneity in the data, we clustered all single cells. We first combined all Drop-seq samples and normalized the data (21,566 cells, 10,791 protein-coding genes detected in at least 3 cells and mean UMI at least 0.005) using regularized negative binomial regression as outlined above (correcting for sequencing depth related factors and cell cycle). We identified 731 highly variable genes; that is, genes for which the *z*-scored standard deviation was at least 1. We used the variable genes to perform dimensionality reduction using diffusion maps as outlined above (with relative eigenvalue cutoff of 2%), which returned 10 significant dimensions. For clustering we used a modularity optimization algorithm that finds community structure in the data with Jaccard similarities (neighbourhood size 9, Euclidean distance in diffusion map coordinates) as edge weights between cells³⁸. With the goal of overclustering the data to identify rare populations, the small neighbourhood size resulted in 15 clusters, of which two were clearly separated from the rest and expressed marker genes expected from contaminating cells (*Neurod6* from excitatory neurons, *Igf1bp7* from epithelial cells). These cells represent rare cellular contaminants in the original sample (2.6% and 1%), and were excluded from further analysis, leaving 20,788 cells.

Identifying a maturation trajectory. To assign each cell a maturation score that is proportional to the developmental progress, we first performed dimensionality reduction as described above using all genes that were detected in at least 2% of the cells (8,014 genes). This resulted in four significant dimensions. We then fit a principal curve (R package princurve, smoother = ‘lowess’, $f = 1/3$) through the data. The maturation score of a cell is then the arc-length from the beginning of the curve to the point at which the cell projects onto the curve. The resulting curve is directionless, so we assign the ‘beginning’ of the curve so that the expression of *Nes* is negatively correlated with maturation. *Nes* is a known ventricular zone marker and therefore should only be highly expressed early in the trajectory. Maturation scores are normalized to the interval [0, 1]. In an independent analysis, we also used Monocle2 to order cells along a pseudo-time. We used Monocle version 2.3.6 with expression response variable set to negative binomial. We estimated size factors and dispersion using the default functions. For ordering cells, we reduced the set of genes based on results of the monocle dispersion Table function, and only considered 718 genes with mean expression ≥ 0.01 and an empirical dispersion at least twice as large as the fitted dispersion. Dimensionality reduction was carried out using the default method (DDRTree).

Defining mitotic and post mitotic populations. We observed a sharp transition point along the maturation trajectory at which cells uniformly transitioned into a postmitotic state, corresponding to the loss of proliferation potential and exit from the cell cycle (Fig. 1f, Extended Data Fig. 1). We therefore subdivided the maturation trajectory into a mitotic and postmitotic phase to facilitate downstream analyses. We defined cells with a high phase-specific enrichment score (score > 2 , see section ‘Removal of cell cycle effect’) as being in the S or the G2/M phase. We then fitted a smooth curve (loess, span = 0.33, degree = 2) to number of cells in S, G2/M phases as a function of maturation score. The point where this curve falls below half the global average marks the dividing threshold (Fig. 1f).

Smoothed expression for visualization. Although all statistical analyses (differential expression, branch detection, etc.) were performed on single-cell data (UMI counts or

normalized expression), we created smoothed expression estimates for visualization in Fig. 2b, c, and generated these by first fitting a loess curve (span = 0.5, degree = 2) to the normalized expression of each gene with maturation score as predictor. We then predicted values at 100 points at regular intervals from 0 to 1.

Identifying developmentally regulated genes. To identify genes that are developmentally regulated during the mitotic phase, we used mutual information between expression and maturation score.

We selected all cells that we determined to be mitotic (as defined above, Fig. 1f, Extended Data Fig. 1) but ignored the 1% of cells with the lowest maturation score to be more robust against outliers. We discretized maturation scores by placing each cell into one of 13 equal sized bins, and did the same for the expression of each gene. We then calculated mutual information between each gene and the maturation score. We also calculated a random background mutual information (rbMI) distribution for each gene using shuffled maturation scores. These rbMI values allowed us to z-score the mutual information values by subtracting the mean of rbMI and then dividing by the standard deviation of rbMI. We determined 1,294 genes to be highly developmentally regulated ($z\text{-score} > 20$).

We call 840 of these genes highly conserved, because all eminence-specific expression fits have a Pearson correlation > 0.9 with the fitted values obtained using the combined data (Fig. 2b).

Differential expression. We wanted to identify genes that were differentially expressed in the early mitotic cells between the eminences (Fig. 2a, Supplementary Table 3), and also between cells assigned to different branches (Fig. 3e, Supplementary Table 4). As has previously been observed³⁹, expression values in scRNA-seq are overdispersed, and we model expression values as drawn from a negative binomial distribution. Concordant with our model for data normalization, our test is based on the same negative binomial regression model with regularized overdispersion parameter. For a gene i and its vector of UMI counts c_i and a group indicator variable g , we fit the two models:

$$\text{Model 1: } \log(E(c_i)) = \alpha + \beta_1 r + \beta_2 m + \beta_3 c + \varepsilon$$

$$\text{Model 2: } \log(E(c_i)) = \alpha + \beta_1 r + \beta_2 m + \beta_3 c + \beta_4 g + \varepsilon$$

with the technical factors r , the total number of reads per cell, and m , the average number of molecules per gene per cell, and the biological factor c , the cell cycle score. The overdispersion parameter θ is determined using model 1 and used for both models. The comparison of the two models using a likelihood ratio test determines the P value of model 2 providing a better fit. The log-fold change is directly given by the coefficient of the group indicator variable as $\log_2(e^{\beta_4})$. We called genes differentially expressed if the adjusted P value (false discovery rate, FDR) is smaller than 10^{-4} and the absolute fold change is larger than 1.

Branch analysis. To check for emerging heterogeneity in the cells, we focused on the postmitotic cells and performed a trajectory analysis that allows for branching, that is, one population of cells may give rise to multiple precursors. Minimum spanning trees (MSTs) have been used previously to identify putative branching structures in developing populations^{18,40}. However, spurious edges in the MST or similar graph structures, previously referred to as short circuits, can introduce stochasticity into these analyses. To overcome this problem, we applied a bootstrapped approach using an ensemble of graphs, an approach inspired by the Wanderlust algorithm⁴¹, which also constructs an ensemble of graphs to gain robustness to short circuits. We repeatedly constructed MSTs based on subsamples of the data, and combined their results to obtain a new cell-to-cell distance matrix and final tree structure that connects all cells. Branches are determined by traversing the final tree and identifying major splits.

We performed this analysis separately for each eminence, and input were the expression data of cells that we considered postmitotic based on maturation score and cell cycle score (1,992 CGE cells, 1,750 LGE cells, 1,271 MGE cells). For the analysis, we normalized the expression data as described above, regressing out the following factors: number of reads, molecules per gene, sample and cell cycle score. We considered all genes that were detected in at least 2% of the cells and that were highly variable, that is, had a z-scored standard deviation larger than 1 (997 genes in CGE, 954 in LGE, 1,017 in MGE). We carried out dimensionality reduction as outlined above. We performed multiple runs of constructing an MST using a random set of 66% of all cells each time. The process was repeated until each pair of cells has been sampled at least 30 times (number of bootstraps for CGE, LGE, MGE: 91, 89, 89). We combined the MSTs by averaging the cell-to-cell distances along the tree structures, followed by MDS to two dimensions. In theory, the MDS was not necessary, but in practice we observed a more robust final consensus MST on the MDS coordinates than on the averaged cell-to-cell distances. The consensus MST is given directionality by choosing a cell with a low maturation score as the root. Instead of using the cell with the lowest maturation score, an approach that can

be sensitive to outliers, we chose the cell that maximizes the correlation between maturation score and distance from the root for all cells.

To determine significant branches, we traversed the tree starting from the root. Any cell that has outgoing edges to two or more cells is a potential branch point. However, we considered branches significant only when the number of cells in the branch was at least 8% of all cells in the MST. We marked only cells that are connected to two or more significant branches as branch points. Only terminal branches, that is, tree segments not containing branch points, are considered for further analysis.

Mapping E14.5 fate-mapped 10x scRNA-seq data to branches. *Lhx6-GFP* mice were crossed with wild-type Swiss Webster females. At E14.5, the MGE and CGE were each dissected from transgenic embryos and dissociated into a single-cell suspension. FACS sorting was used to collect GFP-positive cells from the MGE and GFP-negative cells from the CGE. Single-cell libraries were prepared using the 10x Genomics pipeline. A total of 12,513 cells passed our initial filtering (5,998 *Lhx6*-negative and 6,515 *Lhx6*-positive cells).

Since *Lhx6* is expressed in postmitotic precursors, the *Lhx6*-positive sample from the MGE dissection should contain only postmitotic MGE precursors. However, the *Lhx6*-negative sample derived from the CGE will contain both mitotic progenitors and postmitotic precursors, as well as a minority of cells expressing *Lhx6* mRNA, probably owing to errors during FACS sorting or a time delay in GFP translation. To remove mitotic progenitors from the *Lhx6*-negative CGE dataset, we performed a maturation trajectory analysis as outlined above, and kept only postmitotic cells (2,905 cells, Extended Data Fig. 4h–j). To conservatively remove potential MGE cells from the same dataset, we clustered the data (as described in ‘Initial clustering of all cells’) and removed all cells belonging to clusters with an *Lhx6* detection rate of more than 20%. This step removed 465 cells, leaving 2,440 *Lhx6*-negative postmitotic cells (Extended Data Fig. 4k).

To determine the branch identity of these cells, we mapped them to the E13.5 Drop-seq branches using a correlation-based approach. We focused on 279 genes that were differentially expressed in one branch compared to the other two. For these genes, we averaged the normalized expression in all branches (three branches in each eminence) to create branch model vectors. We then calculated Pearson correlations between all individual cells that we wanted to map and the model vectors (CGE model vectors for the *Lhx6*-negative sample, MGE model vectors for the *Lhx6*-positive sample). We assigned each cell to the branch with the highest correlation, but also calculated empirical P values to determine the significance of the assignment by permuting the single-cell data for a random background. We left the model vectors unchanged, but permuted the single-cell expression data 100 times. For each permutation and each cell we kept track of the largest Pearson correlation to the model vectors, and calculated a P value for the branch assignment by counting what fraction of correlation scores was larger than the one used for the branch assignment. In a final step, we turned all P values into FDRs and mapped only cells with an FDR < 0.1 to the branches (Extended Data Fig. 4l–n).

Mapping E18.5 cortex and subcortex cells to branches. To fate-map *Dlx6a*-lineage neurons, *Dlx6a-cre* mice were crossed with either *Ai9* or *RCE^{loxP}* mice. Cortical and subcortical brain regions were dissected and collected from transgenic embryos at E18.5 and postnatal day (P) 10, dissociated into a single-cell suspension, and cells were collected with FACS based on their fluorescence expression. Single-cell libraries were prepared using the 10x Genomics pipeline. To identify the potential branch of origin of E18.5 cortical (8,382 cells) and subcortical (8,237) neurons, we mapped the cells of the E18.5 samples to the branches using the same approach as for the E13.5 cells. We identified 774 differentially expressed genes between the branches in the E13.5 10x data and used the average across cells as branch model vectors (separately for CGE and MGE derived cells). We then used the same procedure as for the E13.5 cells to map to the branches, allowing each E18.5 cell to map to any of the six E13.5 branches (3 CGE branches, 3 MGE branches), and applying the 0.1 FDR cutoff. Detailed results are shown in Extended Data Fig. 5 and Fig. 3i.

Relative variance explained. To quantify the contribution of different factors to the overall heterogeneity in our data, we compared the amount of variance of individual factors to the variance explained by the first principal component. Given an expression matrix (normalized using negative binomial regression as described above, regressing out number of reads and average number of molecules per detected gene), we first selected the most variable genes ($z\text{-scored standard deviation} > 1$). To quantify the variance associated with an annotated factor of interest, we first constructed a vector representing the annotation of each cell. For a continuous factor of interest (for example, cell cycle score), we centred the vector and length-normalized it to length 1. We then projected the expression data onto this vector, and calculated the variance of the projected dataset. For a discrete factor (for example, branch), we first turned the vector into a set of indicator variables, and applied principal component analysis to obtain independent continuous vectors. We then projected the dataset onto each of these vectors, calculated the

variance of the projected dataset and took the sum over all vectors. This enabled us to compare the variance explained between different annotated factors (for example, cell cycle and maturation score). For visualization and interpretation in Figs. 2e and 3k, we normalized these values by the variance explained by the first principal component of the dataset.

Smart-seq2 data processing. Reads were aligned using the same pipeline as for the Drop-seq data. We kept only cells with at least 2,000 unique genes detected. We further removed cells with a z -scored alignment rate < -3 , or with an absolute z -scored number of reads (after \log_{10} transformation) > 3 . This resulted in 1,099 cells in the Flashtag experiments, and 400 cells in the *Dlx6a*-negative experiments.

Expression data were normalized to x with $x_{i,j} = \log_{10}\left(\frac{c_{i,j} \times 1,000}{m_j} + 1\right)$ in which $c_{i,j}$ is the read count of gene i in cell j and m_j is the sum of all mapped reads for cell j . Cell cycle score was calculated as described above, and regressed out using a linear model including an intercept term. We used the residuals for all further analyses. For dimensionality reduction we used diffusion maps on all genes that were detected in at least 5% of the cells (approximately 10,000 genes). All further analysis was carried out as for the Drop-seq data.

Integration of developmental and adult datasets. We used genetic fate-mapping strategies in combination with the 10x Genomics Chromium system for scRNA-seq to study cortical interneuron development at embryonic (E13.5, E18.5), postnatal (P10) and adult (P56) stages. Specifically, we used the *Lhx6-GFP* transgenic mouse line to select for postmitotic MGE cells and to precisely discriminate MGE versus CGE precursor cells at E14.5 (Extended Data Fig. 4). For later stages, in which cells have migrated out of the ganglionic eminences, we used a *Dlx6a-cre; RCE^{loxP}* pan-ganglionic-eminence fate-mapping strategy to collect cortical interneurons at E18.5 and P10 (Extended Data Fig. 5) and used the publicly available Allen Brain Institute scRNA-seq dataset²² (Allen Cell Types Database; <http://celltypes.brain-map.org/rnaseq>) for the adult time point.

We applied our recently developed integration tool for scRNA-seq datasets²⁵, to identify shared sources of variation between embryonic (E13.5, E18.5), postnatal (P10) and adult (P56) datasets.

For the P56 dataset, we downloaded FPKM expression values for 8,432 single cells from the mouse visual cortex, sequenced with the Smart-Seq2 protocol, from a publicly available resource at the Allen Brain Atlas (Allen Cell Types Database; <http://celltypes.brain-map.org/rnaseq>). We selected 3,432 GABAergic cells for downstream analysis based on an initial clustering analysis and a selection of *Gad1*-positive clusters (Extended Data Fig. 6). For E18.5 and P10 datasets, we performed a similar initial clustering, removing populations of microglia, astrocytes, oligodendrocytes and smooth muscle cells that probably represent FACS false-positives and are unlikely to give rise to cortical interneurons (Extended Data Fig. 5). For the E13.5 and E14.5 dataset, we took all cells from our *Lhx6*-positive and *Lhx6*-negative datasets that were assigned to the interneuron precursor state (branch 1).

We performed three separate pairwise analyses, aligning E13.5 and P56, E18.5 and P56, and P10 and P56 datasets. In each case, we applied the Seurat alignment procedure as previously described²⁵. We first detected variable genes in each dataset independently, using the *FindVariableGenes* function with default parameters. We used the union of the two variable gene sets used as input to canonical correlation analysis, and aligned the resulting canonical correlation vectors (CCV) across datasets with the *AlignSubspace* function. In brief, *AlignSubspace* constructs ‘metagenes’ representing the average expression of genes exhibiting robust correlations to the CCV in both datasets, and applies nonlinear ‘warping’ algorithms to align these metagenes between datasets. We performed this analysis for each of the top 15 CCV independently, and used biweight midcorrelation (*bicor*), a median-based similarity metric implemented in the WGCNA R package⁴². For downstream analysis (*t*-distributed stochastic neighbour embedding (*t*-SNE) and subtype mapping), we selected CCV for which at least 30 genes exhibited a minimum *bicor* of 0.15 in both datasets, applying the same cutoff across all three analyses.

For visualization (Fig. 4b), we constructed a distance matrix from these selected components as input to tSNE with default parameters. We next assigned cells from developmental datasets to adult subtypes, performing the following analysis

separately for each of the three pairwise comparisons. For each cell in the developmental dataset, we calculated the $k = 10$ closest neighbours in the P56 dataset, using the selected CCV for the input distance matrix. If at least 9 of these neighbours in the P56 dataset were of the same subtype, the developmental cell was assigned to this subtype. We note that this represents a stringent threshold, which we apply equally across all comparisons. As a secondary check, for each developmental cell, we calculated the nearest $k = 10$ neighbours across all cells in the merged developmental and adult dataset. If none of these 10 neighbours represented cells in the P56 dataset, we also considered the cell to be unassigned. We performed this mapping procedure twice, to assign cells to the four main cardinal types (Fig. 4b–d), and also to assign them to the 14 finer subtypes (Extended Data Fig. 9).

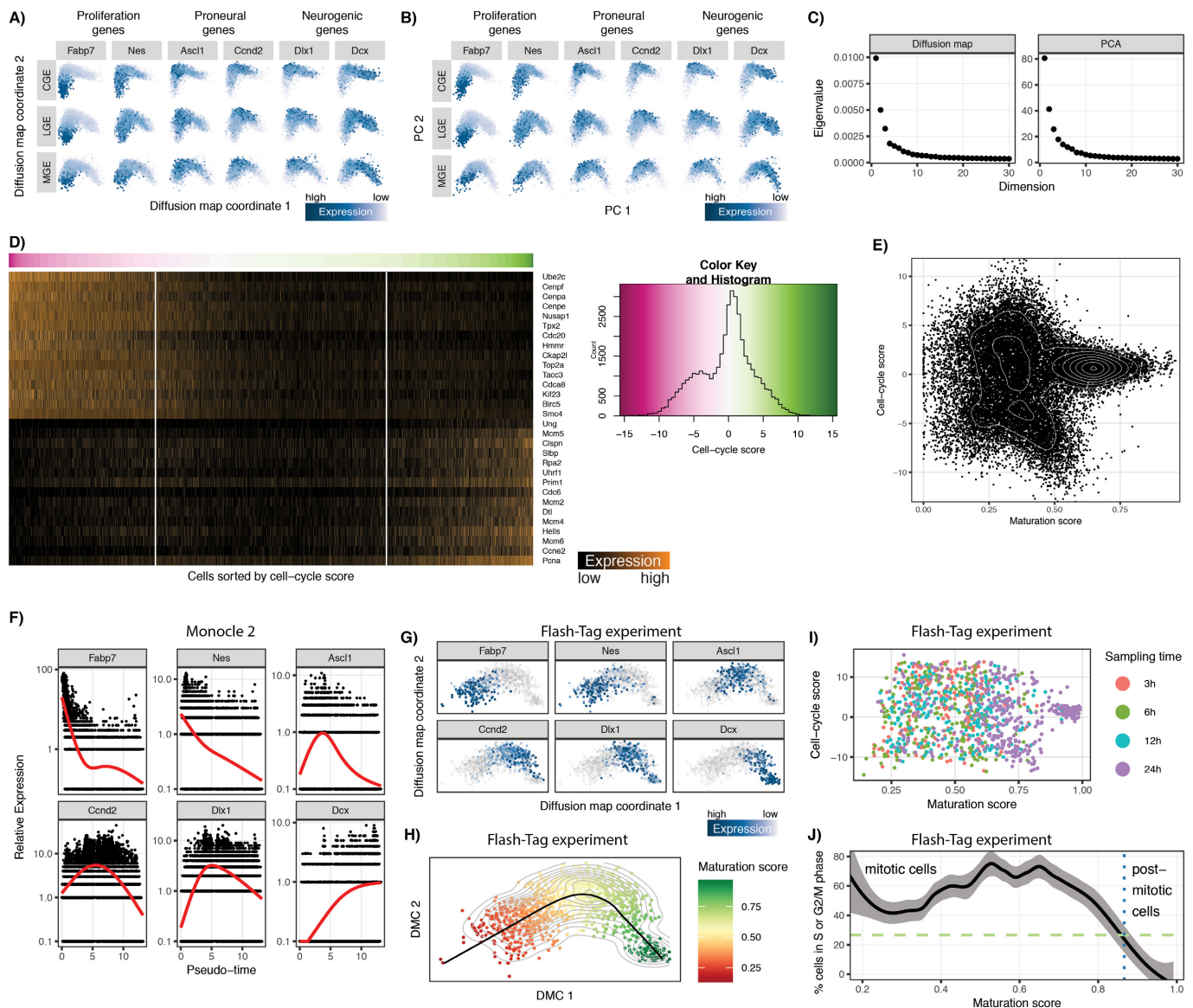
To identify differentially expressed genes that were conserved across development, we used the *FindConservedMarkers* command in Seurat, which runs differential expression tests separately on both developmental and adult datasets. We required genes to have a 1.25-fold change, with a Bonferroni-corrected P value threshold of 0.05 in both developmental and adult datasets to be considered a conserved marker. Figure 4e and Extended Data Fig. 7 lists these genes, in order of when they first become annotated as differentially expressed.

Lastly, we explored whether our early transcriptomic markers of cardinal type separation in the mouse were also differentially expressed in human adult neurons. We used data from a recently published dataset of 14,963 single human nuclei from post-mortem tissue²⁸. We performed standard log-normalization in Seurat, and calculated the average expression level of nuclei that were annotated as belonging to *Pvalb*, *Sst*, *Id2*, and *Vip* types. We used these values for the scatter plots in Extended Data Fig. 10.

Code availability. Code for preprocessing and analysis of scRNA-seq data are available from the Center for Open Science at <https://osf.io/xjmtr>. The Seurat package, which was used for the integration of developmental datasets, is open-source and freely available on GitHub (<https://github.com/satijalab/seurat>) and CRAN (<https://cran.r-project.org/web/packages/Seurat/index.html>).

Data availability. All source data, including sequencing reads and single-cell expression matrices, are available from the Gene Expression Omnibus (GEO) under accession code GSE104158.

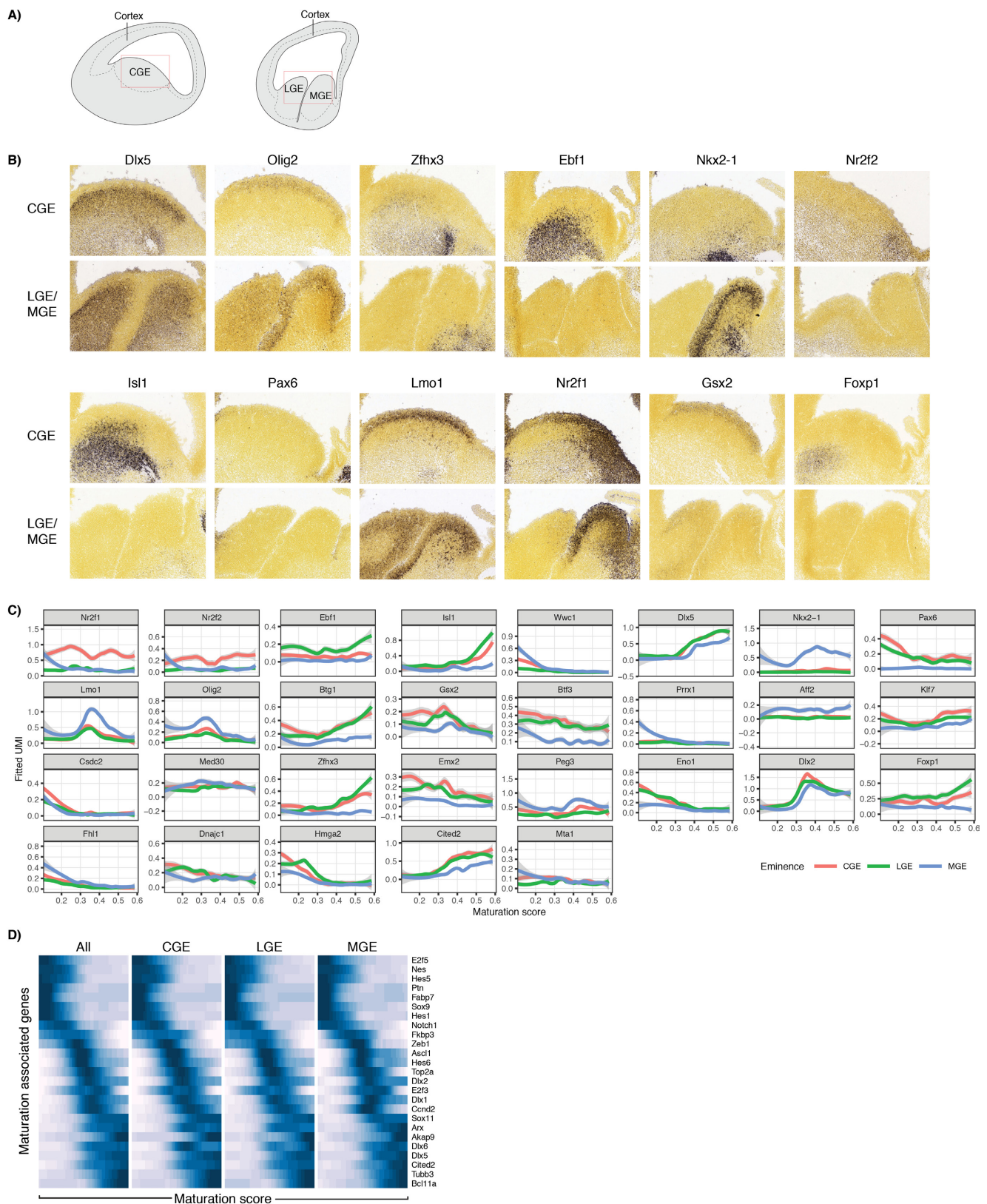
30. Monory, K. *et al.* The endocannabinoid system controls key epileptogenic circuits in the hippocampus. *Neuron* **51**, 455–466 (2006).
31. Madisen, L. *et al.* A robust and high-throughput Cre reporting and characterization system for the whole mouse brain. *Nat. Neurosci.* **13**, 133–140 (2010).
32. Sousa, V. H., Miyoshi, G., Hjerling-Leffler, J., Karayannis, T. & Fishell, G. Characterization of *Nkx6-2*-derived neocortical interneuron lineages. *Cereb. Cortex* **19** (Suppl 1), i1–i10 (2009).
33. Vong, L. H., Ragusa, M. J. & Schwarz, J. J. Generation of conditional *Mef2c^{loxP/loxP}* mice for temporal- and tissue-specific analyses. *Genesis* **43**, 43–48 (2005).
34. Picelli, S. *et al.* Smart-seq2 for sensitive full-length transcriptome profiling in single cells. *Nat. Methods* **10**, 1096–1098 (2013).
35. Lee, S., Hjerling-Leffler, J., Zaghera, E., Fishell, G. & Rudy, B. The largest group of superficial neocortical GABAergic interneurons expresses ionotropic serotonin receptors. *J. Neurosci.* **30**, 16796–16808 (2010).
36. Paxinos, G. & Franklin, K. B. J. *The Mouse Brain in Stereotaxic Coordinates 2nd edn* (Academic, 2003).
37. Haghverdi, L., Büttner, F. & Theis, F. J. Diffusion maps for high-dimensional single-cell analysis of differentiation data. *Bioinformatics* **31**, 2989–2998 (2015).
38. Blondel, V. D., Guillaume, J.-L., Lambiotte, R. & Lefebvre, E. Fast unfolding of communities in large networks. *J. Stat. Mech.* P10008 (2008).
39. Kharchenko, P. V., Silberstein, L. & Scadden, D. T. Bayesian approach to single-cell differential expression analysis. *Nat. Methods* **11**, 740–742 (2014).
40. Qiu, P. *et al.* Extracting a cellular hierarchy from high-dimensional cytometry data with SPADE. *Nat. Biotechnol.* **29**, 886–891 (2011).
41. Bendall, S. C. *et al.* Single-cell trajectory detection uncovers progression and regulatory coordination in human B cell development. *Cell* **157**, 714–725 (2014).
42. Langfelder, P. & Horvath, S. WGCNA: an R package for weighted correlation network analysis. *BMC Bioinformatics* **9**, 559 (2008).



Extended Data Figure 1 | Ordering cells along a maturation trajectory.

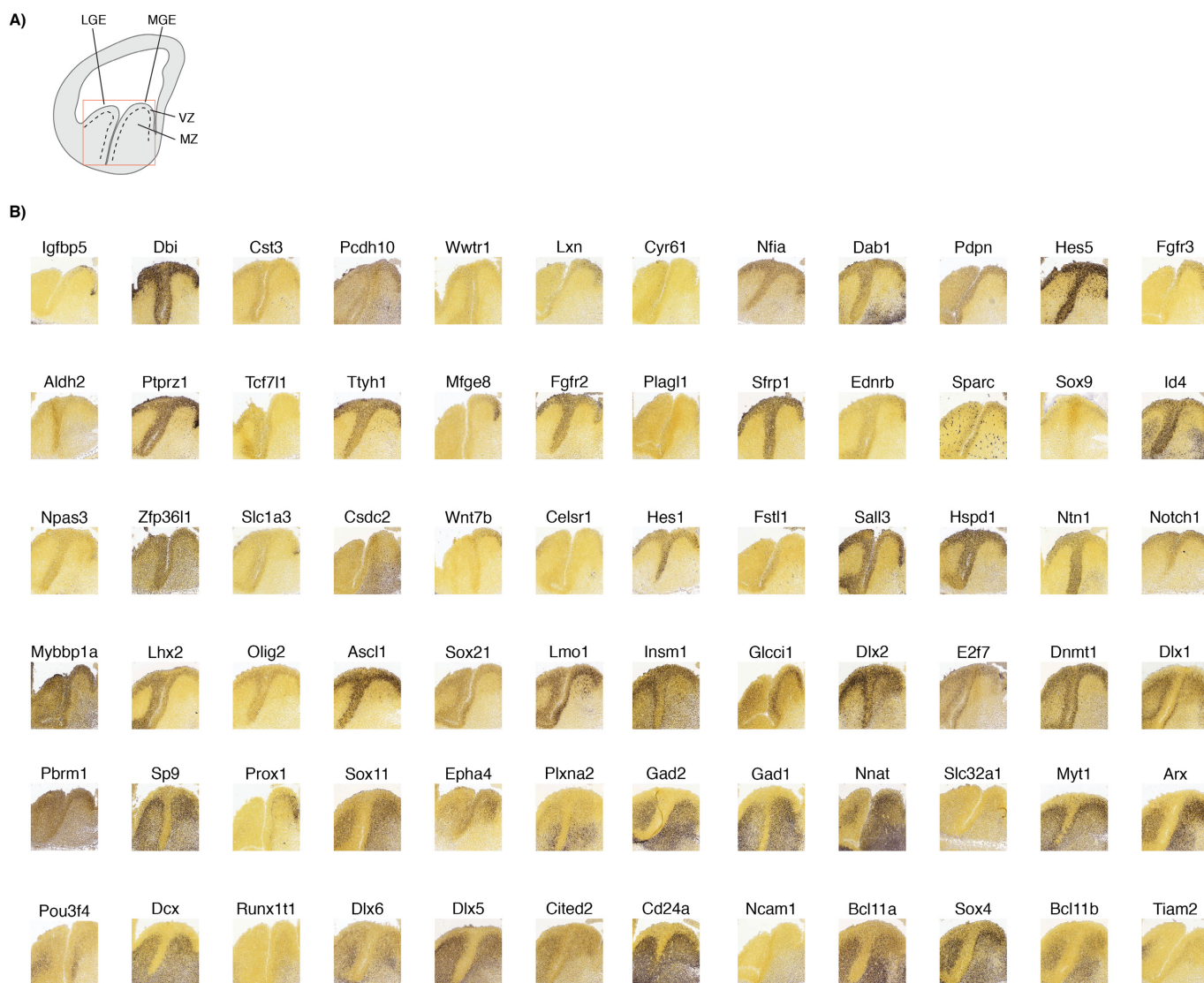
a, Diffusion map analysis of eminence datasets suggests a pan-eminence developmental continuum. Each eminence was analysed independently, revealing nearly identical patterns. Cells are coloured according to the expression of canonical regulators. **b**, Using principal component analysis to reconstruct developmental maturation returns nearly identical results to the diffusion map analysis in Fig. 1. Principal component analysis was calculated for all eminences independently, and cells are coloured by their expression of canonical markers. **c**, Eigenvalues for the two dimensionality reduction methods. We observe a substantial eigenvalue drop-off after the initial components, demonstrating that the majority of the variance is captured in the first few dimensions. **d**, Single-cell heatmap showing scaled expression levels of top genes that were correlated with cell cycle score. Cells on the x axis are sorted by cell cycle score. Negative scores correspond to cells in S phase, positive scores correspond to cells in G2/M phase. **e**, Scatter plot illustrating the relationship between maturation score and cell cycle score for all cells in the dataset. Each dot corresponds

to a single cell. Early progenitors span a wide range of cell cycle states, whereas late cells do not express G2/M or S-phase specific genes and express postmitotic genes. **f**, Expression of canonical marker genes as a function of 'pseudotime', as calculated with Monocle2¹³. Monocle2 pseudotime was strongly correlated with our maturation trajectory (both Pearson and Spearman $R = 0.94$). **g**, **h**, Diffusion map (**g**) and maturation trajectory (**h**) analysis of 1,099 single cells obtained from FlashTag animals, and sequenced using a custom version of the Smart-seq2 protocol (Supplementary Methods). Cells are coloured by their expression of canonical markers, which exhibit dynamics that are concurrent with the maturation trajectory learned from the Drop-seq data. **i**, **j**, Relationship between the maturation trajectory and cell cycle scores derived from the FlashTag datasets replicates our observations from Drop-seq. Therefore, our FlashTag maturation trajectory serves as complementary validation of our Drop-seq maturation trajectory, and exhibits strong association with biological time.



Extended Data Figure 2 | Enrichment of differentially expressed genes in the MGE, CGE and LGE. **a**, Schematic of embryonic brain sections at E13.5/E14.5. One sagittal section shows the MGE and LGE next to one another (right), whereas the other shows the CGE (left). **b**, ISH images from the Allen Brain Institute Developing Mouse Brain Atlas at E13.5 for genes that our analysis identified as being differentially expressed between the eminences. For each gene, ISH images are shown for the MGE, CGE and LGE. Image credit: Allen Institute. **c**, Temporal dynamics

for differentially expressed genes in early mitotic cells. Curves represent local averaging of single-cell expression, as a function of progression along the maturation trajectory, for each eminence independently. Grey area indicates 95% confidence interval. Genes are selected from the differentially expressed genes in early mitotic cells (Fig. 2a). **d**, Gene expression dynamics in mitotic cells, based on local averaging of single-cell data, plotted along maturation score for selected developmentally regulated genes.

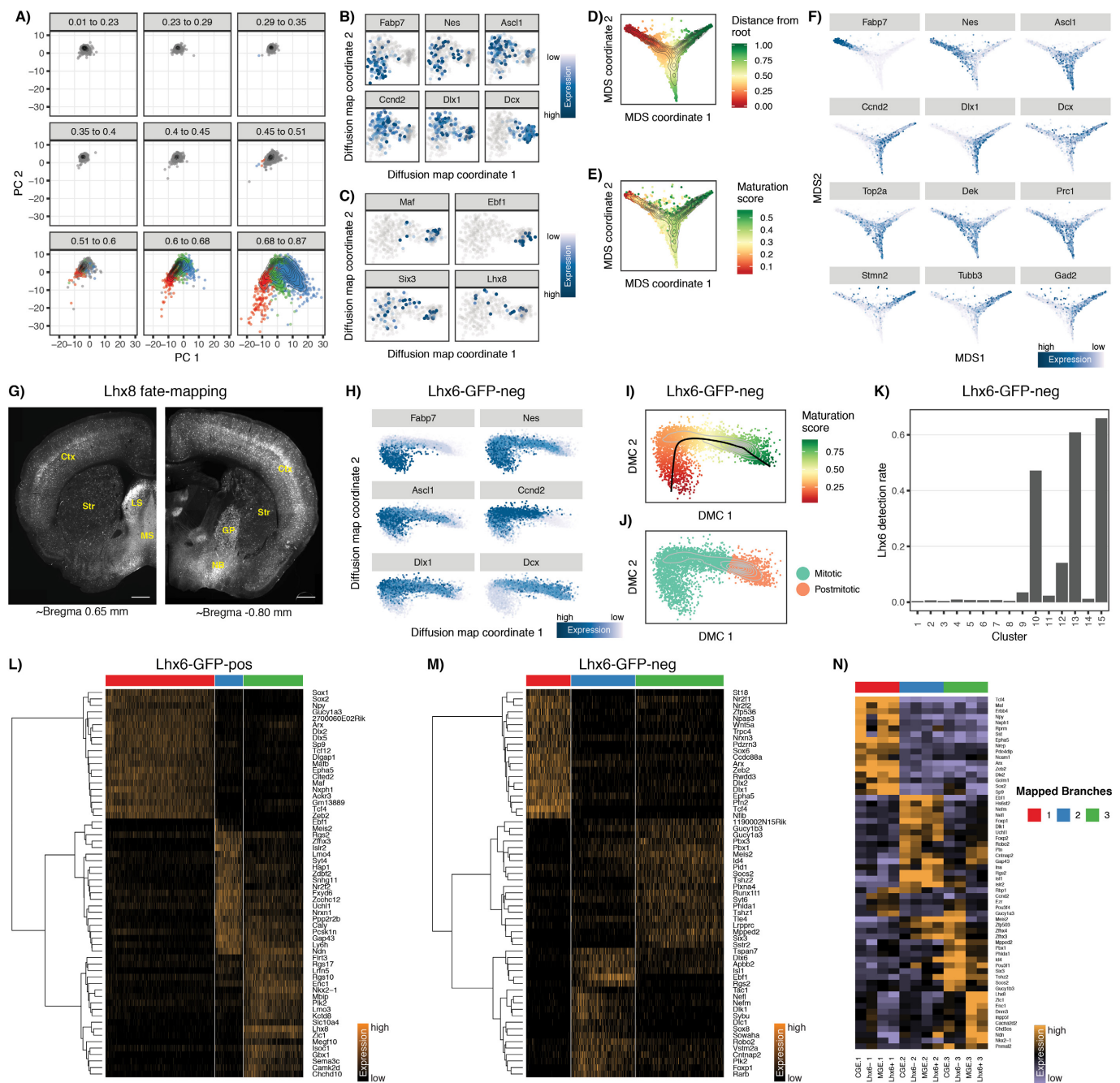


Extended Data Figure 3 | Enrichment of dynamically expressed genes in the ventricular zone, subventricular zone and mantle zone.

a, Schematic of an embryonic brain section at E13.5/E14.5. The locations of the ventricular zone (VZ) and mantle zone (MZ) are indicated.

b, Sagittal ISH images from the Allen Brain Institute Developing Mouse Brain Atlas at E13.5. Genes are ordered from lowest to highest maturation

score rank. The trend overall shows that genes with peak expression at low maturation score tend to have higher expression in the ventricular zone, and as maturation score rank increases the expression pattern shifts to the subventricular zone and then to the mantle zone. Image credit, Allen Institute.



Extended Data Figure 4 | See next page for caption.

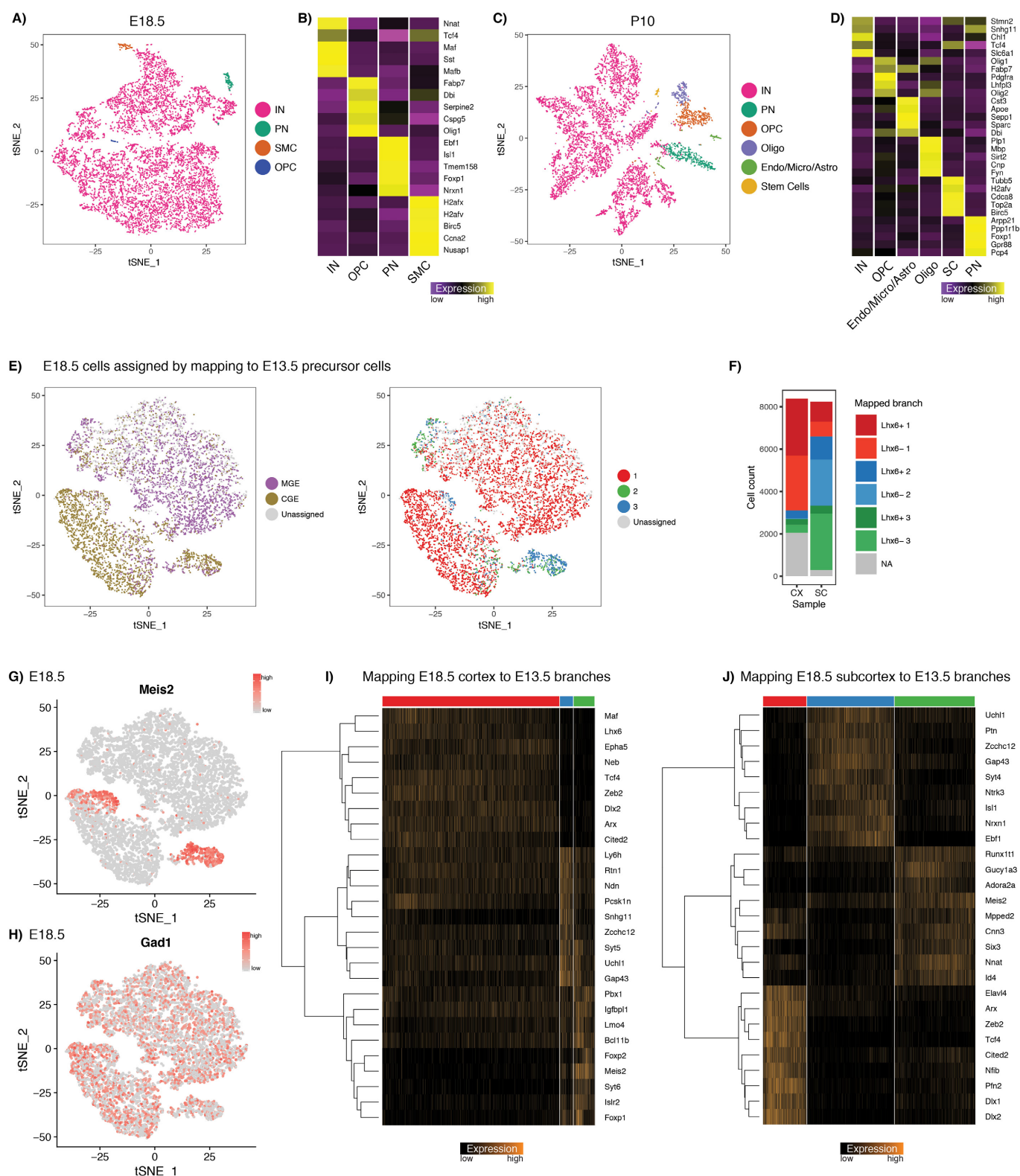
Extended Data Figure 4 | Fate divergence occurs as cells become

postmitotic. a, Supervised analysis: PCA of full dataset, run using only branch-dependent genes. Cells are grouped based on the maturation-trajectory bin: the first five bins represent mitotic progenitors, the last four bins represent postmitotic cells which are coloured by branch ID. Mitotic cells fall within a homogeneous point cloud, with low variance on principal components 1 and 2, showing no evidence of fate bifurcation.

b, To test whether our inability to detect fate bifurcations earlier in development was due to the lower sequencing depth of Drop-seq, we sequenced 400 *Dlx6a-cre;RCE^{loxP}* negative ganglionic eminence cells (thereby enriching for mitotic progenitors), using a modified Smart-seq2 protocol. Diffusion map analysis of these cells returned only two significant principal components, with no evidence of further structure. These components reflect our previously defined maturation trajectory, with DMC1 separating mitotic cells (left). **c**, Rare mitotic cells expressing canonical branch markers do not segregate on the diffusion plot.

d–f, Branching analysis on mitotic progenitors. We repeated the branch analysis, previously computed on postmitotic cells (Fig. 3a), on mitotic progenitors from all three ganglionic eminences. Although we did observe computational evidence of branching, this does not represent fate bifurcations as we observed in postmitotic cells. Instead, cells from different branches could largely be separated into ‘early’, ‘intermediate’ and ‘late’ regions of mitotic pseudotime, with one branch being largely defined by the expression of pro-neural cell cycle regulators (for example, *Ascl1*). As these genes peak at intermediate stages, our branching patterns could reflect either the aberrant assignment of intermediate cells to a new branch, or reflect the potential of multiple modes of cell division (namely,

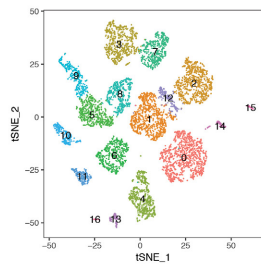
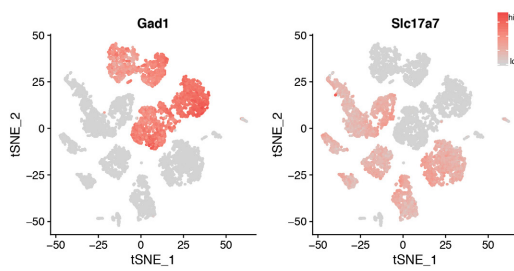
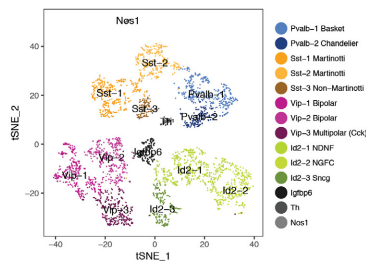
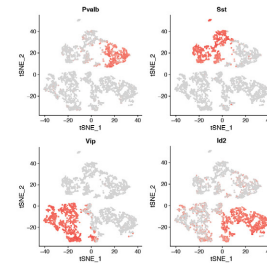
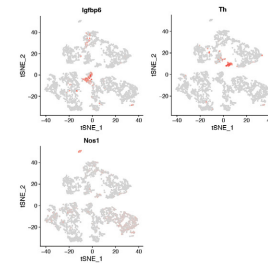
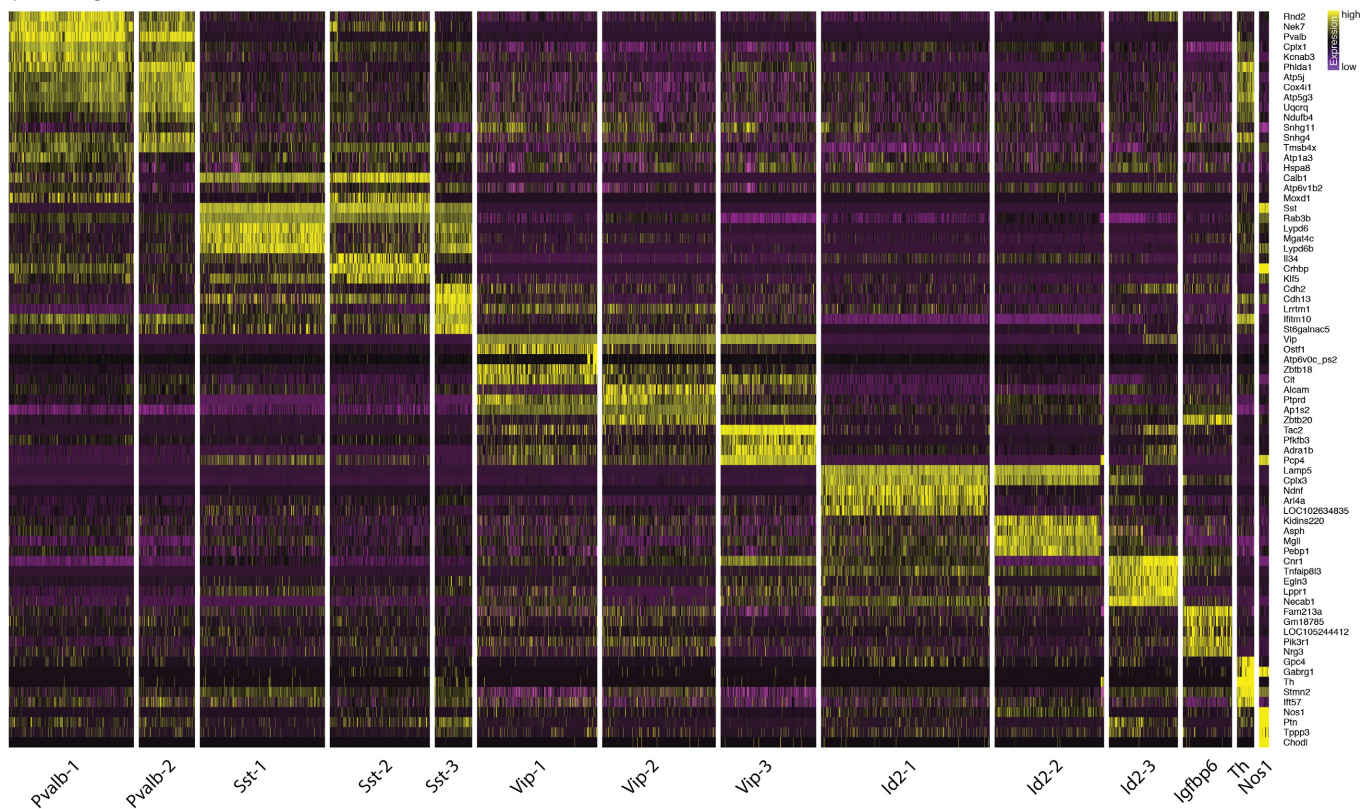
direct versus indirect neurogenesis) occurring in the ventricular zone and subventricular zone. **g**, Genetic fate-mapping using *Lhx8-cre/cerulean* demonstrates that MGE branch three precursors give rise to the entire breadth of cholinergic projection (globus pallidus and nucleus basalis) and interneuron (striatum) populations. The cumulative longitudinal use of a constitutive Cre driver also results in extensive labelling of cortical interneurons owing to transient expression within this population. Scale bar, 500 μ m. Ctx, cortex; Str, striatum; LS, lateral septum; MS, medial septum; NP, nucleus basalis; GP, globus pallidus. **h**, Our *Lhx6*-GFP-negative dataset contains both mitotic and postmitotic cells from the CGE and diffusion map analysis shows our previously defined maturation trajectory. **i, j**, To isolate postmitotic cells, we calculated a maturation trajectory (**i**), and used the cell cycle scores to identify the transition point between mitotic and postmitotic cells (**j**) as with the eminence datasets in Fig. 1. **k**, To avoid the possibility of FACS false-negative MGE cells contaminating our *Lhx6*-GFP-negative dataset, we clustered the postmitotic cells from this dataset, and filtered out three rare clusters where *Lhx6* mRNA expression was detected in more than 20% of cells (Supplementary Methods). **l, m**, We mapped postmitotic cells from the *Lhx6*-GFP-positive (**l**) and *Lhx6*-GFP-negative (**m**) datasets to the branches determined from the Drop-seq dataset (Supplementary Methods). Heat maps show scaled single-cell expression markers associated with each branch. **n**, Analogous to Fig. 3e, but also including the *Lhx6*-GFP-positive and *Lhx6*-GFP-negative datasets generated using 10x Genomics, as a validation of the original Drop-seq datasets that were performed on wild-type mice.



Extended Data Figure 5 | See next page for caption.

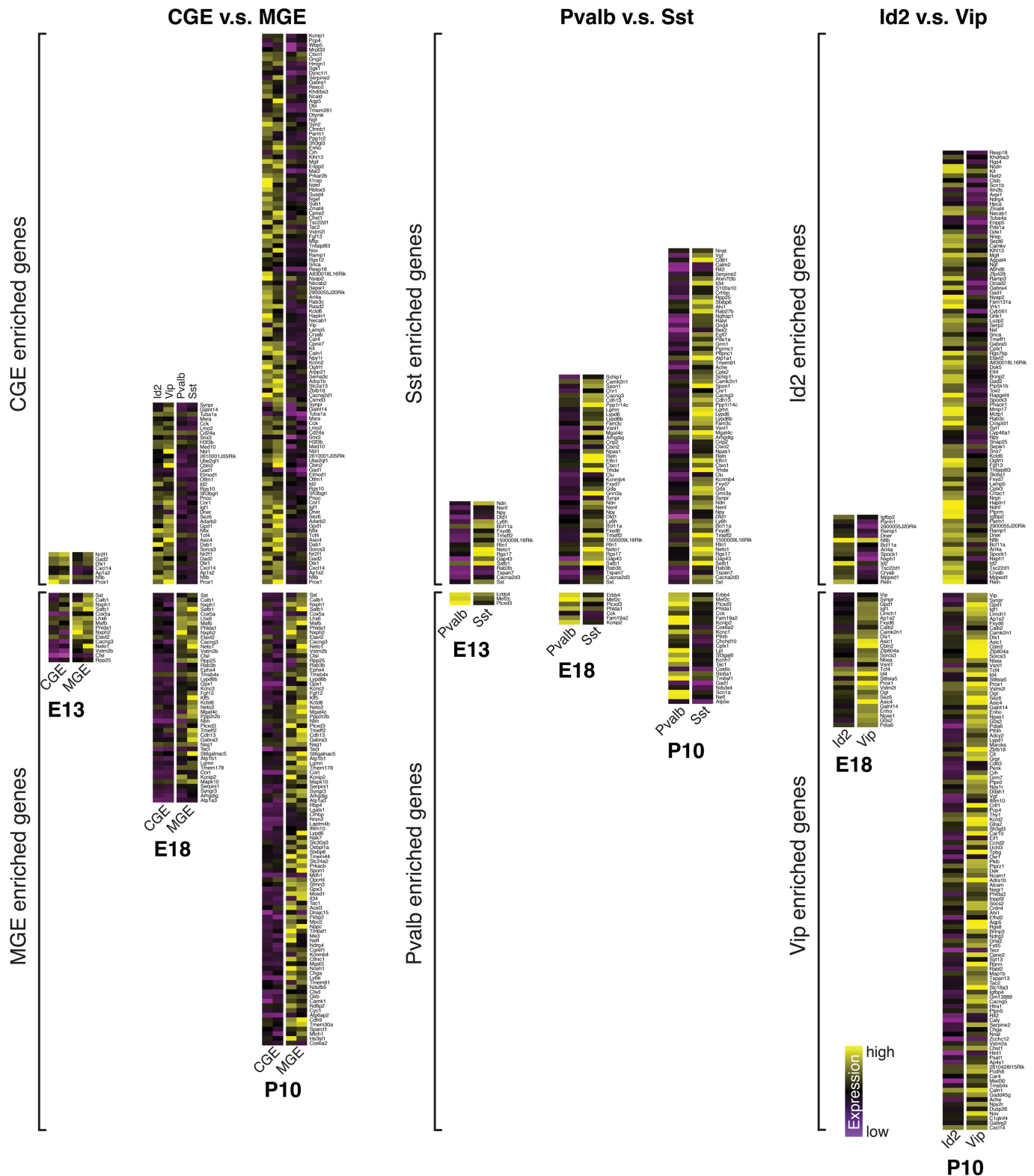
Extended Data Figure 5 | Filtering of E18.5 and P10 10x datasets and mapping of E18.5 cortex and subcortex neurons to E13.5/E14.5 branches. **a, c,** *t*-SNE visualization of *Dlx6a-cre;RCE^{loxP}* positive E18.5 cortical cells (**a**) and *Dlx6a-cre;RCE^{loxP}* positive P10 cortical cells (**c**). Although the *Dlx6a-cre* should mark only GABAergic eminence-derived cells, we identified rare populations that did not express canonical interneuron (IN) markers, probably representing false positives from FACS. **b, d,** Gene expression in these populations (E18.5 cells **b**, P10 cells **d**; heat map shows average expression in group) identifies rare contaminating populations of microglia (micro), astrocytes (astro), oligodendrocyte precursor cells (OPCs) and oligodendrocytes (oligo); smooth muscle cells (SMC), stem cells (SC), projection neurons (PN). For all downstream analyses, we considered only cells in the interneuron cluster. **e,** *t*-SNE visualization of 8,382 *Dlx6a-cre;RCE^{loxP}* positive E18.5 cortical cells (same dataset as in Extended Data Fig. 5a, but after removing contaminating populations). Each E18.5 cell was mapped to one of six precursor states (branch 1, 2, and 3 for Lhx6–GFP-positive and

Lhx6–GFP-negative datasets), using a correlation-based distance metric (Supplementary Methods). This enabled us to assign a putative eminence and branch of origin for each of the E18.5 cortical cells. **f,** As expected, the vast majority of *Dlx6a-cre;RCE^{loxP}* positive E18.5 cortical cells map to the interneuron precursor state, and are split between MGE and CGE-derived precursors. By contrast, *Dlx6a-cre;RCE^{loxP}* positive E18.5 cells from the subcortex primarily map to branches 2 and 3, consistent with our interpretation of these branches as precursor states for projection neurons; CX, cortex; SC, subcortex. **g, h,** The minority of *Dlx6a*-positive cortical cells mapping to precursor states 2 and 3 primarily co-express *Meis2* (**g**) and *Gad1* (**h**), probably representing a CGE-derived GABAergic population. These cells have been recently described as being present in the cortical white matter and probably represent projection neuron precursors²¹. **i, j,** Heat maps showing single-cell expression markers for the three different mapped branches of *Dlx6a-cre;RCE^{loxP}* positive E18.5 cells from the cortex (**i**) and the subcortex (**j**).

A) Adult Visual Cortex (Data from the Allen Institute)**B)****C) GABAergic neurons in the adult visual cortex****D)****E)****F) GABAergic neurons in the adult visual cortex****Extended Data Figure 6 | Clustering of adult visual cortical neurons into 14 major non-overlapping inhibitory interneuron subtypes.**

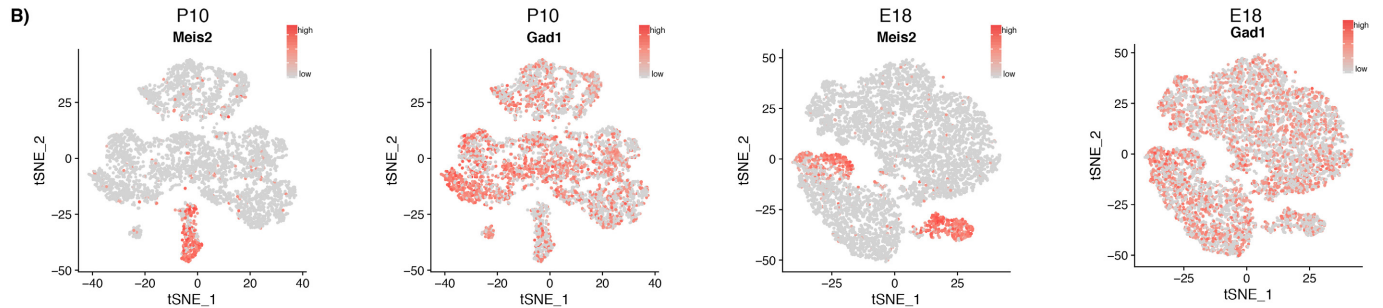
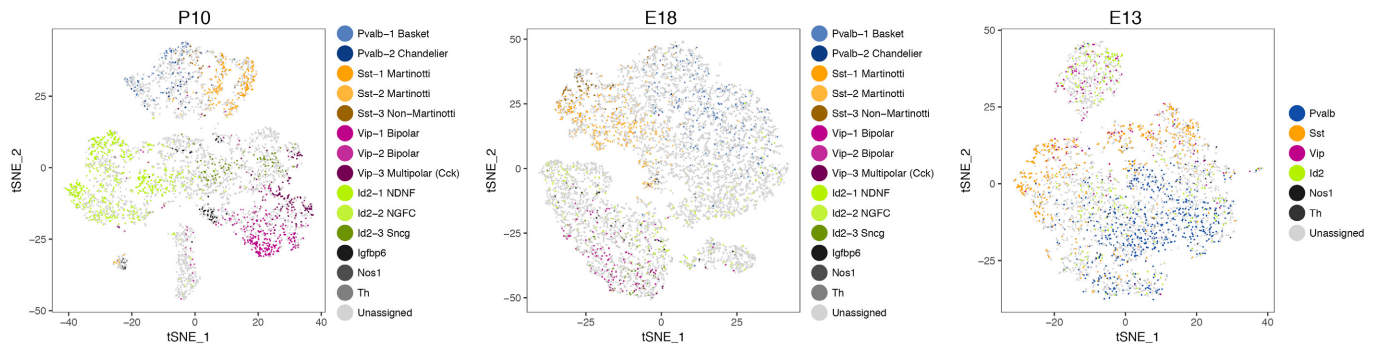
a, Initial *t*-SNE visualization and graph-based clustering of 8,329 single cells individually isolated from P56 mouse visual cortex and sequenced with the Smart-Seq2 protocol. Data was downloaded from the publicly available resource hosted by the Allen Brain Atlas²² (Allen Cell Types Database, <http://celltypes.brain-map.org/download> (2015)). **b**, Of all cells,

3,432 GABAergic interneurons were easily identified by the expression of *Gad1* (left) and the absence of *Slc17a7* (right), and were selected for downstream analysis. **c**, *t*-SNE visualization and graph-based clustering of the 3,432 GABAergic cells reveals 14 clusters. **d**, **e**, The clusters revealed in **c** could be broadly grouped into cardinal types based on the expression of canonical markers. **f**, Single-cell heat map showing scaled expression values for the best transcriptomic markers in each cluster.



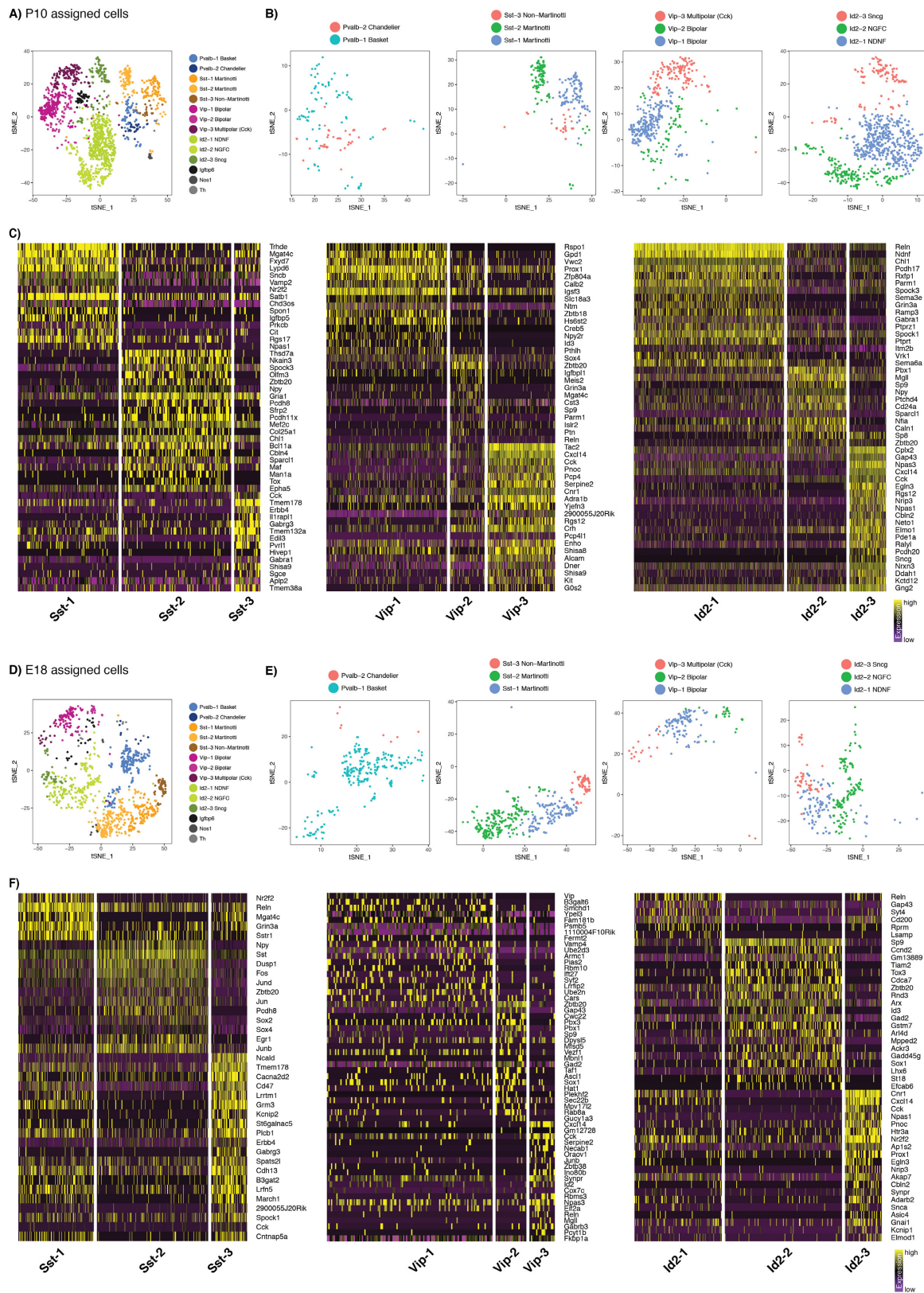
Extended Data Figure 7 | Emergence of transcriptomically defined subtypes across development. Differentially expressed genes between MGE and CGE derived subsets (left), that are conserved in both developmental and P56 cells. Each conserved gene is placed on the heat map when it is first observed to be differentially expressed during

development, and the number of conserved differentially expressed genes grows over time. The same analysis is shown for Pvalb and Sst subsets (middle), and for Vip and Id2 subsets (right). This figure is identical to Fig. 4e, but with all gene names displayed.

A) Cells assigned by mapping to P56 neurons

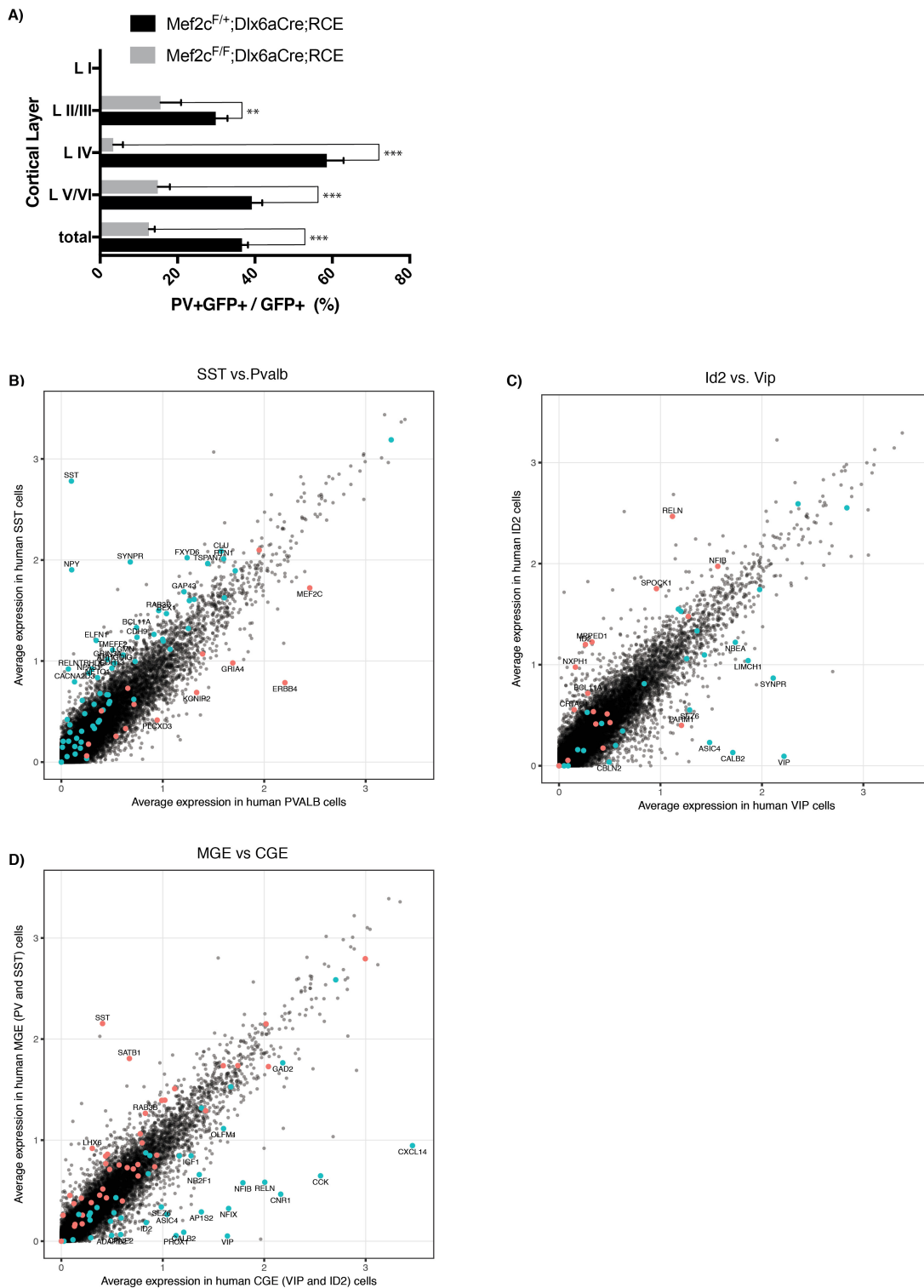
Extended Data Figure 8 | The integrated analysis agrees with an independent *t*-SNE analysis of each time point. **a**, *t*-SNE visualizations of interneuron precursors from E13.5, E18.5 and P10, calculated independently for each time point. Cells are coloured as in Fig. 4b–d, based on their mapping to P56 datasets in integrated analysis. However, since the *t*-SNE was performed separately for each time point, we can assess how the integrated analysis agrees with an independent analysis of each time point. In each case, we can see that the cardinal type separation

that we observe via integrated analysis (Fig. 4b–d) is consistent with an independent analysis of each dataset. Integrated analysis with the P56 dataset results in clearer separation, and enables us to map developmental precursors to adult subtypes. **b**, Expression of *Gad1* and *Meis2* in single-cell datasets. Cells expressing both genes are probably projection neuron precursors that have recently been described in the CGE²¹, but whose progeny is not captured in the mouse visual cortex dataset. Therefore, these cells are correctly mapped as unassigned.



Extended Data Figure 9 | Transcriptional segregation into cortical interneuron subtypes at different developmental stages. a, *t*-SNE visualization of all P10 cells mapping to a P56 subtype (as in right column of Fig. 4c, but cells are coloured by subtype instead of cardinal type). b, *t*-SNE visualization as in a, but zoomed in on each cardinal type independently. c, Single-cell heat maps showing the best transcriptomic markers marking each subtype, for the Sst (left), Vip (middle) and Id2 (right) cardinal types, within P10 cells. We did not observe any statistically

significant markers subdividing Pvalb subtypes. d, *t*-SNE visualization of all E18.5 cells mapping to a P56 subtype (as in right column of Fig. 4c, but cells are coloured by subtype instead of cardinal type). e, *t*-SNE visualization as in d, but zoomed in on each cardinal type independently. f, Single-cell heat maps showing the best transcriptomic markers marking each subtype, for the Sst (left), Vip (middle) and Id2 (right) cardinal types, within E18.5 cells. We did not observe any statistically significant markers subdividing Pvalb subtypes.



Extended Data Figure 10 | A subset of embryonic markers of cardinal type specification in mouse are conserved in adult human neurons.

a, Quantification of Pvalb-positive cortical interneurons across the different cortical layers of the control and *Mef2c* cortical knockout (*Dlx6a-cre; Mef2c^{loxP/loxP}RCE*) animals. *Mef2c* cortical knockout results in a reduction in Pvalb density in all cortical layers except for layer 1. Error bars reflect s.e.m.; unpaired *t*-test; **P* < 0.05, ***P* < 0.01, ****P* < 0.001; *n* = 4 brains each for cortical knockout and control, 3–4 sections per

brain. **b–d**, Scatter plot comparing average expression of 3,035 GABAergic single nuclei from post-mortem human neurons, after segregation into Pvalb and Sst (**b**), Vip and Id2 (**c**) and MGE and CGE inferred origins (**d**). Each dot represents the expression of a gene in human cells. Markers of transcriptomic cardinal types from our E13.5 and E18.5 datasets (from Fig. 4e) are shown in red or blue dots. Mouse embryonic markers that also differ by 1.5-fold in human have gene names annotated on the plot.

Placentation defects are highly prevalent in embryonic lethal mouse mutants

Vicente Perez-Garcia^{1,2*}, Elena Fineberg^{1,2*}, Robert Wilson³, Alexander Murray^{1,2}, Cecilia Icoresi Mazzeo⁴, Catherine Tudor⁴, Arnold Sienerth^{1,2}, Jacqueline K. White⁴, Elizabeth Tuck⁴, Edward J. Ryder⁴, Diane Gleeson⁴, Emma Siragher⁴, Hannah Wardle-Jones⁴, Nicole Staudt⁴, Neha Wali⁴, John Collins⁴, Stefan Geyer⁵, Elisabeth M. Busch-Nentwich^{4,6}, Antonella Galli⁴, James C. Smith³, Elizabeth Robertson⁷, David J. Adams⁴, Wolfgang J. Weninger⁵, Timothy Mohun³ & Myriam Hemberger^{1,2}

Large-scale phenotyping efforts have demonstrated that approximately 25–30% of mouse gene knockouts cause intrauterine lethality. Analysis of these mutants has largely focused on the embryo and not the placenta, despite the crucial role of this extraembryonic organ for developmental progression. Here we screened 103 embryonic lethal and sub-viable mouse knockout lines from the Deciphering the Mechanisms of Developmental Disorders program for placental phenotypes. We found that 68% of knockout lines that are lethal at or after mid-gestation exhibited placental dysmorphologies. Early lethality (embryonic days 9.5–14.5) is almost always associated with severe placental malformations. Placental defects correlate strongly with abnormal brain, heart and vascular development. Analysis of mutant trophoblast stem cells and conditional knockouts suggests that a considerable number of factors that cause embryonic lethality when ablated have primary gene function in trophoblast cells. Our data highlight the hugely under-appreciated importance of placental defects in contributing to abnormal embryo development and suggest key molecular nodes that govern placenta formation.

Systematic identification of the genes required for normal embryogenesis is essential if we are to successfully unravel the molecular framework that underpins embryo development. Such knowledge will help to identify genetic causes of developmental abnormalities that manifest during pregnancy or at birth, and comprise a considerable health burden. Large-scale phenotyping efforts have consistently found that 25–30% of mouse gene knockouts result in non-viable offspring^{1–5}. In almost all studies of developmentally critical genes, research has focused on the impact of the mutation on the embryo. By contrast, little attention has been paid to the possible effects of these mutations on extraembryonic tissues, almost certainly resulting in underrepresentation of placental phenotypes in public databases. The Mouse Genome Informatics (MGI) database, for example, shows extraembryonic defects in only 10% of embryonic-lethal strains. Gaining a more accurate view of the actual frequency of placental abnormalities is crucially important for our understanding of the contribution of this vital organ to the aetiology of developmental defects and congenital abnormalities⁶.

Several groundbreaking studies have highlighted the essential role of extraembryonic tissues for normal development and long-term health. Placental insufficiency results in intrauterine growth retardation and, as a consequence, can cause fetal programming effects that predispose to later-onset disease^{7,8}. Moreover, tetraploid complementation⁹ or conditional gene ablation experiments have identified embryonic-lethal phenotypes in which normal development can be entirely rescued solely by providing the embryo with a wild-type placenta^{10–15}.

However, systematic efforts to discover the genes required for normal placental development are still missing. The Deciphering the Mechanisms of Developmental Disorders (DMDD; <https://dmdd.org.uk>) consortium¹⁶ is one of several ongoing programmes dedicated to

identifying and characterizing embryonic-lethal genes in the mouse. In addition to detailed phenotypic assessment of structural abnormalities in mutant embryos, the DMDD project also investigates the effect of each mutation on placental development. Here we report the analysis of placental morphology for 103 such lines. Our results reveal a markedly higher rate of placental phenotypes than had been previously appreciated, and a striking association of placental defects with specific abnormalities in the embryo itself. Our study identifies the placenta as a pivotal target organ for the effects of gene mutations that contribute to developmental demise.

Placental defects in embryonic-lethal mouse lines

We analysed 103 mouse knockout lines that fail to produce mutant offspring at the expected Mendelian frequency at postnatal day (P)14, but yielded mutant embryos at either embryonic day (E)14.5 or E9.5. Lines for which mutant conceptuses could not be recovered at E9.5 were not included in this screen. Of the 103 lines analysed, 82 were classified as P14 lethal, as no mutant offspring were recovered at that stage. The remaining 21 lines were termed sub-viable, with mutant pups constituting 13% or less of all offspring obtained, a proportion significantly below the 25% expected from heterozygous crosses (Fig. 1a, Supplementary Table 1). Similar criteria were applied to sub-categorise the P14 lethal group further according to viability at E14.5 (Fig. 1a).

Placentas of all lines were subjected to histopathological analysis at E9.5, E14.5 or both (Supplementary Table 1). As expected, less than 1% of wild-type placentas showed an abnormal phenotype. By contrast, in mutant placentas we detected dysmorphologies in 56 out of 82 (68%) of P14 lethal strains (Fig. 1b). Even when the P14 sub-viable lines were

¹The Babraham Institute, Babraham Research Campus, Cambridge CB22 3AT, UK. ²Centre for Trophoblast Research, University of Cambridge, Downing Street, Cambridge CB2 3EG, UK. ³The Francis Crick Institute, 1 Midland Road, London NW1 1AT, UK. ⁴Wellcome Trust Sanger Institute, Cambridge CB10 1SA, UK. ⁵Division of Anatomy, Center for Anatomy & Cell Biology, Medical University of Vienna, Waehringerstrasse 13, A-1090 Vienna, Austria. ⁶Department of Medicine, University of Cambridge, Cambridge CB2 0QQ, UK. ⁷Sir William Dunn School of Pathology, University of Oxford, Oxford OX1 3RE, UK.

*These authors contributed equally to this work.

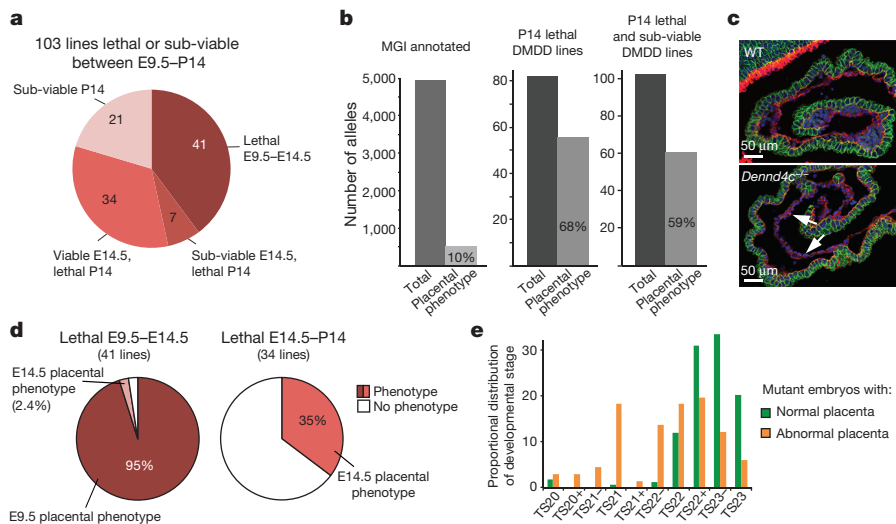


Figure 1 | Placental defects are highly prevalent in gene mutants that affect embryonic viability.

a, Summary of the 103 mouse lines screened. ‘Sub-viable’ identifies strains in which the proportion of mutant offspring is $>0\%$ but $\leq 13\%$. **b**, Summary of non-viable mouse lines in which a placental phenotype has been annotated in the MGI database (<http://www.informatics.jax.org>) and in our DMDD programme. **c**, Yolk sac appearance in wild-type (WT) and *Dennd4c* mutants. Images are representative of three independent mutants and more than 60 wild-type samples analysed. Sections were stained for E-cadherin (green, demarcating the visceral endoderm) and laminin (red, highlighting the basement membrane). Arrows denote the disconnected mesoderm and endoderm layers in mutants. **d**, Breakdown of the proportion of placental phenotypes by stage of embryonic lethality. **e**, Developmental progression of E14.5 mutant embryos depending on the presence or absence of a placental phenotype. TS, Theiler stage.

included, the placental phenotype rate was still 59%, a far higher frequency than the approximately 10% annotated in the MGI database (Fig. 1b). All genes associated with placental abnormalities in mutants were expressed in the trophoblast lineage of this organ (Extended Data Fig. 1a), lending support to the notion that they contribute directly to placental growth or function.

We also assessed the conceptuses for yolk sac defects, an extra-embryonic structure that is especially important for nutrient provision during the first half of gestation, before formation of the functional placenta at E9.5 (Supplementary Table 1). Because the yolk sac was routinely used for genotyping, it proved only possible to analyse this tissue in 66 lines. Among these, an abnormal yolk sac morphology was detected in 11% (7 out of 66) of cases, compared to approximately 6% annotated among prenatal lethals in the MGI database. Notably, all seven affected lines fell within the E9.5–E14.5 lethal group (7 out of 24 = 29%; Fig. 1c). Thus, although yolk sac defects that affect its structure or haematopoietic function may contribute to the lethality of some of these early lethal strains, they occur at a much lower frequency than placental abnormalities.

When scoring the occurrence of placental defects as a function of developmental stage, we found that almost every line that died before E14.5 exhibited placental abnormalities (40 out of 41; Fig. 1d, Extended Data Fig. 1b), compared to only 35% of lines that were viable beyond E14.5 (12 out of 34) (Fig. 1d, Supplementary Table 1). These findings demonstrate that mutations resulting in embryonic lethality between E9.5 and E14.5 are almost certainly associated with a defective placenta.

In line with the placenta being the essential nutrient-supplying organ from mid-gestation onwards, we also found that mutant E14.5 embryos in strains exhibiting a placental phenotype were shifted to a younger developmental stage compared to those in which placental development was normal¹⁷ (Fig. 1e, Extended Data Fig. 1c).

Categories of placental defects

To categorise the different types of defect, we examined the three main layers of the mature placenta: the labyrinth, which constitutes the main nutrient and gas-exchange surface; the junctional zone, which consists of spongiotrophoblast, glycogen cells and different giant cell subtypes; and the maternally derived decidua (Fig. 2, Extended Data Fig. 2a).

Haematoxylin and eosin histology (Extended Data Fig. 2b, c) was complemented with three histological staining methods to accurately classify the cellular and tissue composition defects in abnormal placentas using a series of phenotype criteria (Fig. 2a, b). At E9.5, a frequently detected malformation affected the invagination of allantoic blood vessels into the chorionic ectoderm, a process crucial for the development

of the labyrinth that will almost certainly result in developmental arrest (Fig. 2a, c, Extended Data Fig. 2b). At E14.5, by far the most prevalent abnormalities were defects in the growth and intricate organization of the fetal and maternal blood conduits within the labyrinth layer (Fig. 2b, d, Extended Data Fig. 2c). Because these abnormalities diminish the surface area available for nutrient transport, they will compromise fetal growth and survival.

Collectively, these histological characterizations of more than 300 mutant placentas provide a vast resource for the research community (all data are available at <https://dmdd.org.uk>).

Critical nodes in placental development

We next examined whether the identity of genes associated with placental defects suggested specific molecular pathways that may be pivotal for the formation or function of this organ. For genes that affect placental morphology at E9.5 in mutants, this network analysis highlighted several functional gene clusters that centred around *L3mbtl2*, *Bap1* and *Arhgef7* (Fig. 2e, Extended Data Fig. 1d). Similarly, several factors identified in the E14.5 analysis formed specific molecular nodes, for example, around *Traf2*, *Nek9* and *Rpgrip11* (Extended Data Fig. 1d). Although relatively few genes have been analysed for defects in extraembryonic tissues in the literature, it is obvious that a large fraction of network components identified in our analyses have been associated with embryonic phenotypes. It therefore seems highly likely that mutants for many of these functionally connected genes will also exhibit placental abnormalities.

Embryo and placenta defects are linked

Because the DMDD program scores both embryonic and placental defects, it provides a unique opportunity to assess co-associations between specific phenotypes^{18,19}. Importantly, DMDD phenotype calls are based on precise embryo sub-staging, and the analysis therefore excludes any apparent phenotypes that simply reflect the developmental delay prevalent among embryos with placental defects. Nevertheless, mutant mouse lines that exhibit placental abnormalities were enriched for specific E14.5 embryo phenotypes that were distinct from those with normal placentas (Extended Data Fig. 3a).

Embryo phenotype categories that show significant statistical correlation with placental defects included abnormalities in the heart, brain and vascular system (Fig. 3a, Extended Data Figs 3b, 4a, b and Supplementary Table 2). In particular, this affected anomalies in fore-brain development, heart chamber and septum morphology, subcutaneous oedema, and overall artery or vein topology (Fig. 3b–d, Extended Data Fig. 4c, d). These phenotype co-associations suggest co-regulatory or inter-dependent mechanisms during the development of particular

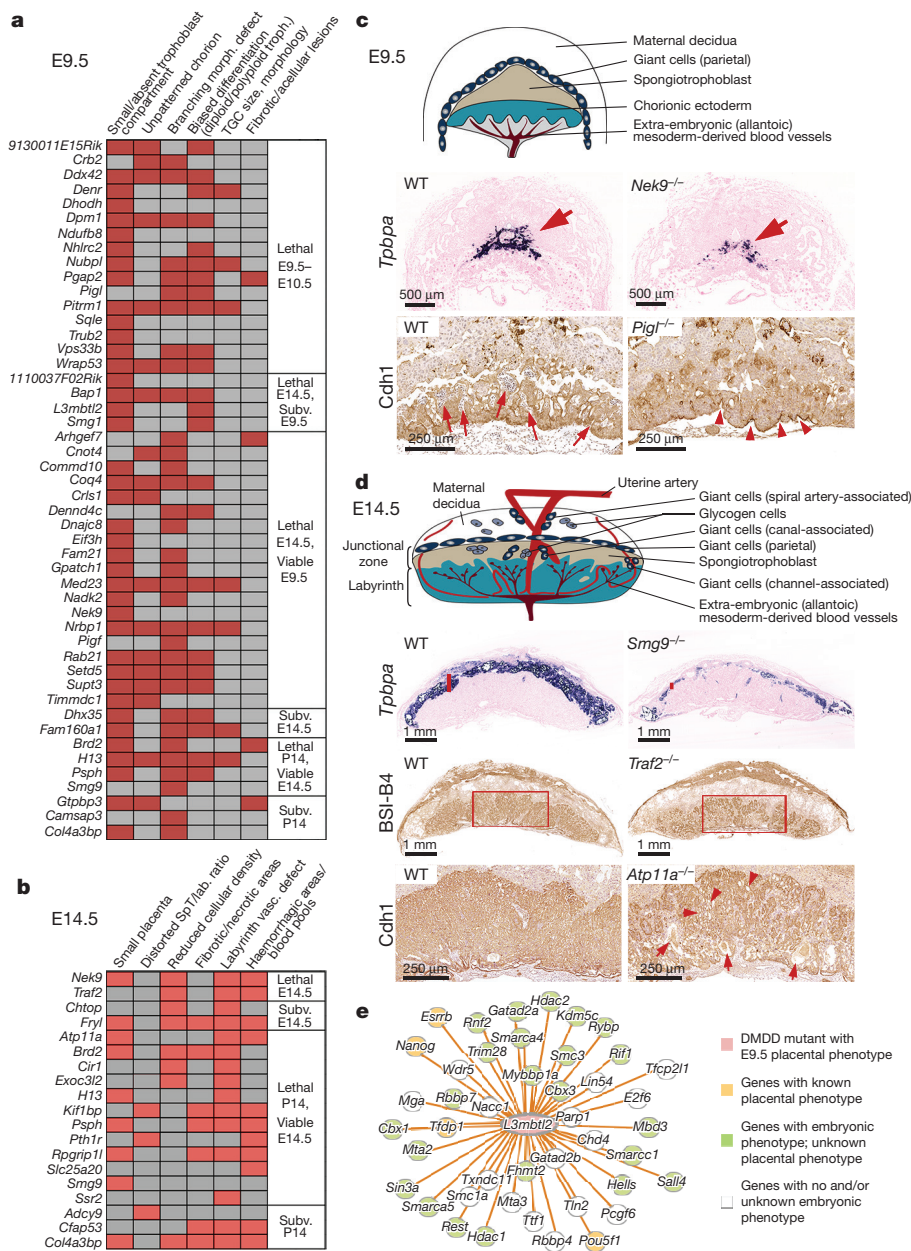


Figure 2 | Summary of common placental defects and functional networks. **a**, Common phenotype criteria used to assess E9.5 mutant placentas (red denotes abnormality detected). Morph., morphology; Subv., sub-viable; TGC, trophoblast giant cell. 1110037F02Rik is also known as *Virma*; *Fam21* is also known as *Washc2*. **b**, E14.5 placental phenotypes in mutant strains. SpT, spongiorophoblast; lab., labyrinth. **c**, Top, schematic representation of main structures of an E9.5 placenta. Bottom, *in situ* hybridization for spongiorophoblast marker *Tbpba* and immunostaining against E-cadherin (*Cdh1*) on wild-type and mutant placentas, as indicated. Large red arrows highlight *Tbpba*-positive cells. Small red arrows in the *Cdh1*-stained wild-type placenta highlight nucleated blood cells in fetal blood vessels. Arrowheads in the *Pigf*^{-/-} placenta demarcate sites of chorionic ectoderm invagination but absence of blood vessels. **d**, Top, schematic representation of the main structures of an E14.5 placenta. Bottom, examples of histological analyses of E14.5 wild-type and mutant placentas, as indicated. Red vertical lines in the *Tbpba* *in situ* hybridization images show thickness of the junctional zone. The BSI-B4 isolectin stainings demarcate the three main placental layers, and the red rectangle highlights the severely reduced complexity of labyrinth vascularisation in the *Traf2* mutant. Red arrowheads in the *Cdh1* immunohistochemistry images point to widened blood spaces of the *Cdh1*-labelled syncytiotrophoblast; arrows denote fibrotic areas. Images in **c** and **d** are representative of at least three independent mutants per line, see Methods. **e**, Network created using esyN (<http://www.esyn.org>) of known interactors of *L3mbtl2*.

organ systems, notably between the placenta and morphogenesis of the brain, heart and vascular system.

Trophoblast-specific gene functions

Because the placenta comprises cell types of distinct lineage origins, a placental phenotype may be caused by trophoblast-intrinsic and/or extraembryonic mesoderm-derived endothelial cell defects. To determine trophoblast-specific functions of genes identified as important for placental development, we used CRISPR-Cas9-mediated ablation in trophoblast stem cells (TSCs)^{20–22} (Extended Data Fig. 5). We chose three genes for this analysis that caused lethality around E9.5–E10.5 when ablated: the tumour suppressor *Bap1*, the cell polarity complex member *Crb2* and the nucleotide binding protein-like factor *Nubpl* (Extended Data Fig. 6).

Nubpl-mutant TSCs exhibited a decreased stem-cell potential, as evidenced by lower expression levels of *Cdx2*, *Esrrb* and *Elf5*, which may explain the marked reduction in the size of the trophoblast compartment in *Nubpl*^{-/-} placentas. Moreover, severely impaired upregulation of *Gcm1* (an early marker of syncytializing trophoblast) and lower expression levels of *Syna* and, to a lesser extent, *Synb* showed

that differentiation towards the syncytiotrophoblast lineage was inhibited in the absence of *Nubpl* (Fig. 4a, Extended Data Fig. 7a). We also detected a prominent phenotype in *Bap1*-deficient TSCs, as they displayed increased expression of the key stem-cell markers *Cdx2* and *Esrrb* when grown under self-renewal conditions. When triggered to differentiate, *Bap1*^{-/-} TSCs failed to upregulate markers of syncytiotrophoblast, sinusoidal trophoblast giant cells and glycogen cells (Fig. 4b, Extended Data Fig. 7b). These TSC differentiation defects may well contribute to the labyrinth formation phenotype evident in both *Nubpl* and *Bap1* mutants. By contrast, *Crb2*-null TSCs were indistinguishable from wild-type (empty vector) controls (Extended Data Fig. 7c).

Lineage origins of placental defects

To gain further insights into trophoblast-intrinsic versus embryonic lineage-induced effects, we chose the same three genes that we studied in TSCs for conditional gene ablation *in vivo*. Thus, we used the *Sox2-cre* transgene to delete the function of these genes in the embryo, while leaving expression intact in the trophoblast-derived cells of the placenta and the visceral yolk sac endoderm²³ (Fig. 5a).

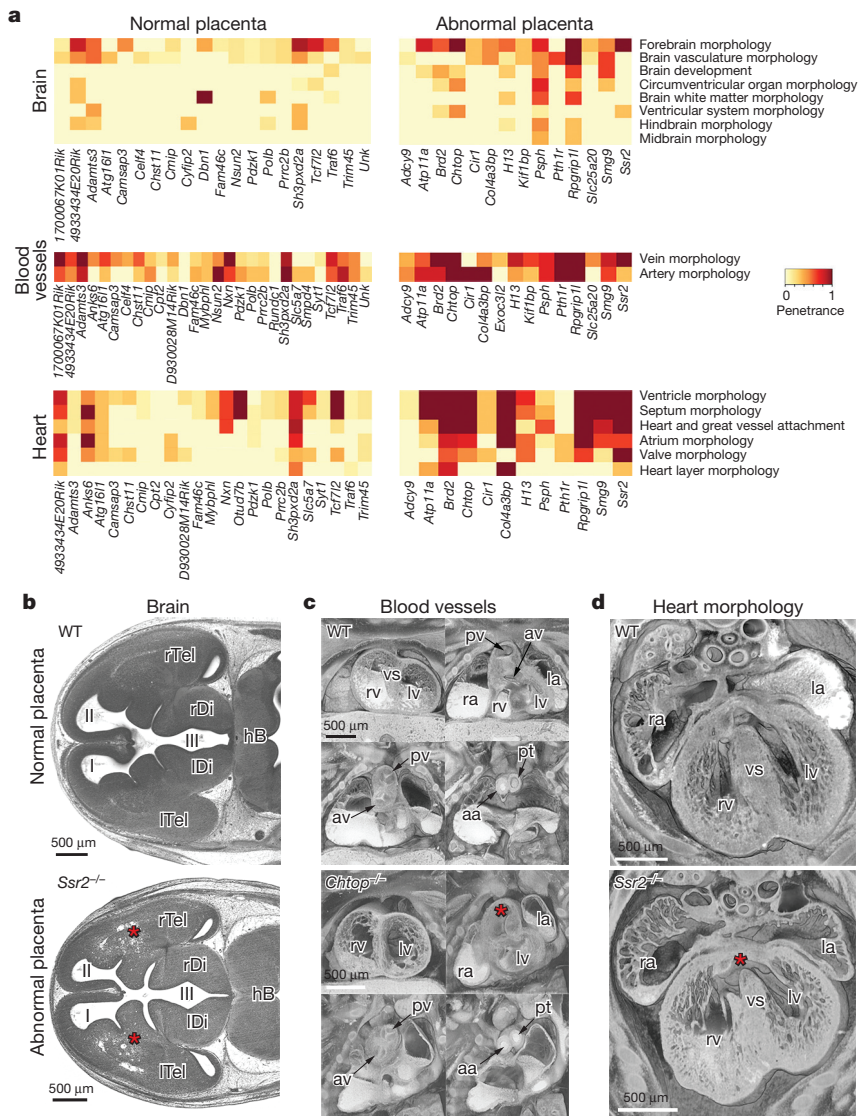


Figure 3 | Phenotype co-associations between embryo and placenta. **a**, Enriched embryonic phenotype terms within the significantly co-associated categories of abnormal brain, blood vessel and heart morphology in mutant lines with abnormal placentas, compared to those with normal placentas (dark red: fully penetrant phenotype). For brevity, the description 'abnormal' has been removed from ontology terms. The most prevalent terms describing abnormalities observed in brain, blood vessel and heart development are shown. **b–d**, High-resolution episcopic microscopy (HREM) images showing embryonic phenotypes that correlate with the presence of placental defects. Top, normal morphology in stage-matched controls. Bottom, distinct developmental abnormalities in corresponding structures of mutants. **a**, Abnormal forebrain morphology (asterisks) in an *Ssr2*^{-/-} embryo. **c**, Double outlet right ventricle and bicuspid aortic valve in a *Chlop*^{-/-} embryo. Right ventricle (rv) with oblique outlet (asterisk). **d**, Perimembraneous ventricular septal defect (asterisk) in an *Ssr2*^{-/-} embryo. I, II and III, first, second and third ventricle, respectively. aa, ascending aorta; av, aortic valve; hb, hindbrain; la/ra, left/right atrial appendix; lDi/rDi, left/right diencephalon; lTel/rTel: left/right telencephalon; lv/rv: left/right ventricle; pt, pulmonary trunk; pv, pulmonary valve; vs, ventricle septum. Defects in **b–d** are representative of at least three independent mutants.

Nubpl-null embryos associated with a heterozygous placenta were considerably more advanced in development than their complete knockout counterparts at E9.5 and could still be recovered up to E11.5, a stage when the complete knockout was already resorbed (Fig. 5a, Extended Data Fig. 8a). Histological examination of the E9.5 and E11.5 placentas showed that the trophoblast expansion, syncytiotrophoblast differentiation and labyrinth vascularization defects were seemingly fully rescued in the conditional knockouts (Fig. 5b, Extended Data Fig. 8b). This rescue was also suggested by the transcriptome-wide similarity between conditional knockout and wild-type or heterozygous control placentas (Extended Data Fig. 9a). Thus, although the conditional *Nubpl* mutation is still lethal beyond E11.5 owing to an essential role of this gene in the embryo proper, a functional trophoblast lineage rescues the placental phenotype and the resulting early mid-gestation embryonic lethality.

For *Bap1*, syncytiotrophoblast formation was partially restored in conditional knockout placentas. Furthermore, global expression profiles of *Bap1* conditional knockout placentas were more similar to controls than to knockouts (Extended Data Fig. 9a, b). However, placental vascularization remained underdeveloped and the conceptuses still died at mid-gestation (Extended Data Fig. 9b). This indicates an essential additional function of *Bap1* in the extraembryonic mesoderm compartment that prevents placental labyrinth formation and also results in a yolk sac defect in knockouts and conditional knockouts (Extended Data Fig. 10a). Similarly, *Crb2*-null embryos could not be rescued by a

genetically functional trophoblast lineage (Extended Data Fig. 10b), a result consistent with the lack of phenotype in mutant TSCs. Because the yolk sac phenotype also remained unchanged in conditional knockouts, it can be concluded that the chorio-allantoic placental defect is due to the crucial role of *Crb2* in mesoderm development²⁴.

In summary, *in vitro* and *in vivo* analysis of three genes whose mutation causes mid-gestational lethality identified two factors (*Nubpl* and *Bap1*) with important roles in the proper expansion and differentiation capacity of trophoblast cells. One of these (*Nubpl*) is indeed causative of the embryonic lethal phenotype at E9.5.

Discussion

Systematic mouse knockout phenotyping efforts undertaken so far have excluded the analysis of extraembryonic tissues, most notably the placenta^{3–5,25}. Ignoring placental defects as a major contributory factor to fetal demise has previously led to several prominent examples of misannotation of gene function, such as for the tumour suppressor *Rb1* and the oncogene *Myc*^{26–29}. In both cases, subsequent studies have revealed that restoring gene function to the trophoblast lineage could largely rescue the embryonic defects observed^{30,31}. Here, we report a systematic effort to assess the prevalence of placental abnormalities in P14 lethal or sub-viable mouse mutants that survive to at least mid-gestation.

We find a remarkably high percentage of placental abnormalities among these lines, with two-thirds of all P14 lethal strains exhibiting obvious defects. In particular, knockouts that result in mid-gestational

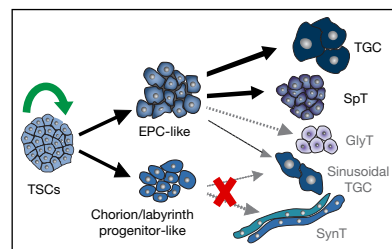
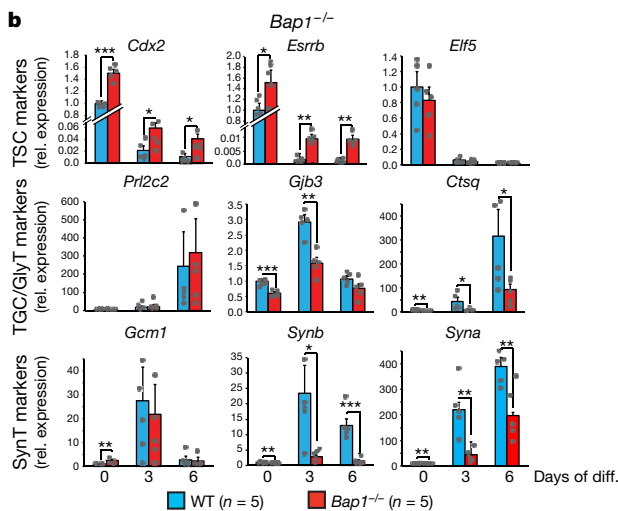
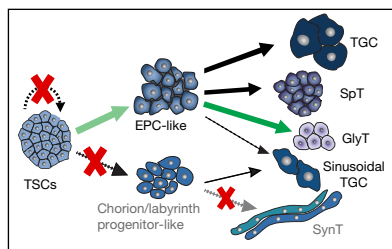
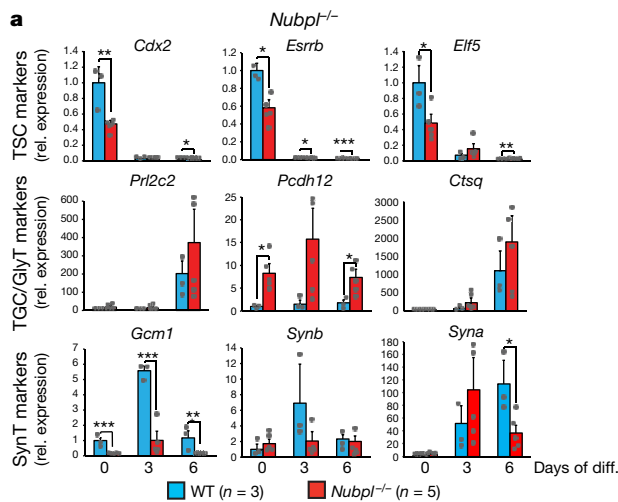


Figure 4 | Determining trophoblast-specific gene function. **a**, Analysis of *Nubpl*^{-/-} TSCs grown in self-renewal conditions (day 0) or after differentiation (diff.) for 3 and 6 days. Data are mean \pm s.e.m. **P* < 0.05; ***P* < 0.01; ****P* < 0.001 (ANOVA with Holm–Bonferroni's post-hoc test) of *n* = 3 (wild-type) and *n* = 5 (*Nubpl*^{-/-}) individual clones as independent replicates. Specific defects are summarized in the schematic. **b**, Equivalent analysis for *Bap1*^{-/-} TSCs of *n* = 5 (wild-type) and *n* = 5 (*Bap1*^{-/-}) individual clones as independent replicates. EPC, ectoplacental cone; GlyT, glycogen cells; rel., relative; SynT, syncytiotrophoblast (layers I and II); TGC, trophoblast giant cells; *Prl2c2* is also known as *Plf*; *Gjb3* is also known as *Cx31*.

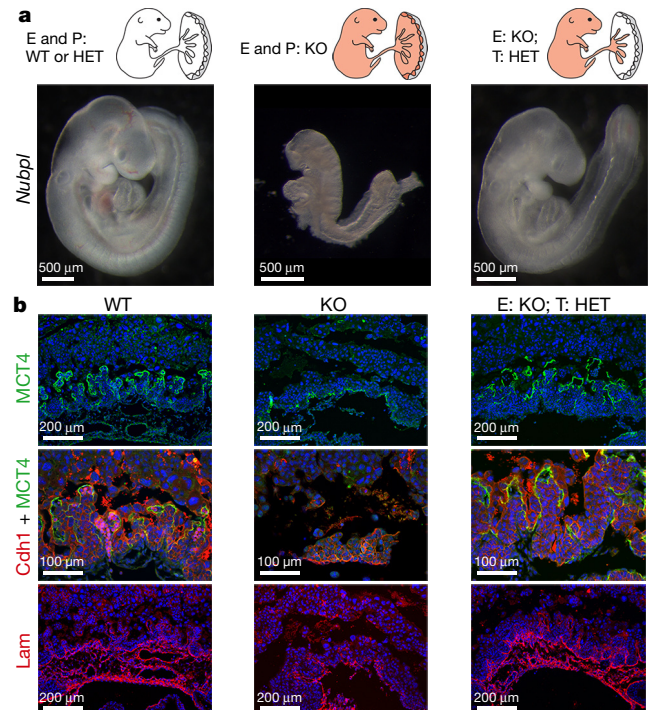


Figure 5 | Dissecting lineage origins of placental phenotypes.

a, Schematic representation of the genetic constitutions of embryo (E) and placenta (P) or trophoblast (T) achieved by conditional *Sox2-cre*-mediated knockout (KO), and corresponding E9.5 embryos of the *Nubpl* strain. Phenotypes are representative of at least 12 embryos per genotype. HET, heterozygous. **b**, Immunofluorescence staining of corresponding placentas for MCT4 (marker of syncytiotrophoblast layer II), E-cadherin and basement membrane component laminin (Lam; demarcates fetal blood vessels). Nuclear counterstain with 4,6-diamidino-2-phenylindole (DAPI). Placental defects are representative of at least three independent mutants per genotype.

lethality are almost certainly associated with an abnormal placenta, underpinning the notion that defects in placentation create a bottleneck for developmental progression past mid-gestation³². This frequency of placental defects illustrates the hugely underestimated effect of gene mutations on extraembryonic tissues. Given that approximately 25–30% of all mutations cause embryonic lethality, our data suggest that a placental phenotype has gone unnoticed and unreported in hundreds if not thousands of mutant strains.

Many of the genes associated with placental defects in our screen are part of specific functional hubs, such as the *L3mbtl2* Polycomb group complex and the tumour necrosis factor-receptor associated factor (*Traf2*) network, which seem to be of major importance for placental development. Identification of such molecular nodes holds great promise as a way of gaining insights into the causes of placental defects in humans. Consistent with this, at least three of the genes we assessed, *TRAF2*, *PSPH* and *BAP1* (through its established interaction with ASXL3) have been implicated in the pathophysiology of human pregnancy disorders, many of which have their origin in defective placentation^{33–36}.

A unique feature of our study is the integrated analysis of both embryo and placenta. This has revealed important co-associations between the occurrence of a placental phenotype and particular defects within the embryo itself, notably affecting neurodevelopment, the heart and the overall vascular system. A placenta–heart axis has been recognized before^{37–40}; however, we can now identify highly specific pathologies such as a double outlet right ventricle and ventricular septal defects that strongly correlate with the presence of an abnormal placenta. Effects of placental insufficiency on brain development have also been reported^{31,41,42}; our large-scale screen provides strong correlative

evidence to support this developmental co-relationship. By contrast, a systematic effect of the placenta on vascular development, beyond overall haemodynamics⁴³, has not previously been recognized. The importance of our findings may therefore extend not only through the immediate gestational period, but also into postnatal life, and may help to explain how placental insufficiency can have long-lasting consequences on cardiovascular disease risk, outweighing other behavioural factors⁴⁴.

Taken together, we demonstrate that placental malformations are far more common than previously thought in embryonic-lethal mutations and co-occur specifically with heart, brain and vascular network defects. Our data highlight the importance of including extraembryonic tissues in studies investigating the genetic basis of congenital abnormalities.

Online Content Methods, along with any additional Extended Data display items and Source Data, are available in the online version of the paper; references unique to these sections appear only in the online paper.

Received 15 February 2017; accepted 23 January 2018.

Published online 14 March 2018.

1. Ayadi, A. *et al.* Mouse large-scale phenotyping initiatives: overview of the European Mouse Disease Clinic (EUMODIC) and of the Wellcome Trust Sanger Institute Mouse Genetics Project. *Mamm. Genome* **23**, 600–610 (2012).
2. de Angelis, M. H. *et al.* Analysis of mammalian gene function through broad-based phenotypic screens across a consortium of mouse clinics. *Nat. Genet.* **47**, 969–978 (2015).
3. White, J. K. *et al.* Genome-wide generation and systematic phenotyping of knockout mice reveals new roles for many genes. *Cell* **154**, 452–464 (2013).
4. Adams, D. *et al.* Bloomsbury report on mouse embryo phenotyping: recommendations from the IMPC workshop on embryonic lethal screening. *Dis. Model. Mech.* **6**, 571–579 (2013).
5. Dickinson, M. E. *et al.* High-throughput discovery of novel developmental phenotypes. *Nature* **537**, 508–514 (2016).
6. Rossant, J. & Cross, J. C. Placental development: lessons from mouse mutants. *Nat. Rev. Genet.* **2**, 538–548 (2001).
7. Barker, D. J. The origins of the developmental origins theory. *J. Intern. Med.* **261**, 412–417 (2007).
8. Barker, D. J., Bull, A. R., Osmond, C. & Simmonds, S. J. Fetal and placental size and risk of hypertension in adult life. *Br. Med. J.* **301**, 259–262 (1990).
9. Rossant, J. Development of the extraembryonic lineages. *Semin. Dev. Biol.* **6**, 237–247 (1995).
10. Guillemot, F., Nagy, A., Auerbach, A., Rossant, J. & Joyner, A. L. Essential role of *Mash-2* in extraembryonic development. *Nature* **371**, 333–336 (1994).
11. Luo, J. *et al.* Placental abnormalities in mouse embryos lacking the orphan nuclear receptor ERR- β . *Nature* **388**, 778–782 (1997).
12. Yamamoto, H. *et al.* Defective trophoblast function in mice with a targeted mutation of *Ets2*. *Genes Dev.* **12**, 1315–1326 (1998).
13. Wang, J., Mager, J., Schnedier, E. & Magnuson, T. The mouse *PcG* gene *eed* is required for *Hox* gene repression and extraembryonic development. *Mamm. Genome* **13**, 493–503 (2002).
14. Shi, W. *et al.* Choroideremia gene product affects trophoblast development and vascularization in mouse extra-embryonic tissues. *Dev. Biol.* **272**, 53–65 (2004).
15. Schreiber, M. *et al.* Placental vascularisation requires the AP-1 component fra1. *Development* **127**, 4937–4948 (2000).
16. Mohun, T. *et al.* Deciphering the Mechanisms of Developmental Disorders (DMDD): a new programme for phenotyping embryonic lethal mice. *Dis. Model. Mech.* **6**, 562–566 (2013).
17. Geyer, S. H. *et al.* A staging system for correct phenotype interpretation of mouse embryos harvested on embryonic day 14 (E14.5). *J. Anat.* **230**, 710–719 (2017).
18. Karp, N. A., Heller, R., Yaacoby, S., White, J. K. & Benjamini, Y. Improving the identification of phenotypic abnormalities and sexual dimorphism in mice when studying rare event categorical characteristics. *Genetics* **205**, 491–501 (2017).
19. Weninger, W. J. *et al.* Phenotyping structural abnormalities in mouse embryos using high-resolution episcopic microscopy. *Dis. Model. Mech.* **7**, 1143–1152 (2014).
20. Tanaka, S., Kunath, T., Hadjantonakis, A. K., Nagy, A. & Rossant, J. Promotion of trophoblast stem cell proliferation by FGF4. *Science* **282**, 2072–2075 (1998).
21. Murray, A., Siennerth, A. R. & Hemberger, M. Plet1 is an epigenetically regulated cell surface protein that provides essential cues to direct trophoblast stem cell differentiation. *Sci. Rep.* **6**, 25112 (2016).
22. Latos, P. A. *et al.* Elf5-centered transcription factor hub controls trophoblast stem cell self-renewal and differentiation through stoichiometry-sensitive shifts in target gene networks. *Genes Dev.* **29**, 2435–2448 (2015).
23. Hayashi, S., Lewis, P., Pevny, L. & McMahon, A. P. Efficient gene modulation in mouse epiblast using a *Sox2Cre* transgenic mouse strain. *Mech. Dev.* **119** (Suppl 1), S97–S101 (2002).
24. Xiao, Z. *et al.* Deficiency in *Crumb* homolog 2 (*Crb2*) affects gastrulation and results in embryonic lethality in mice. *Dev. Dyn.* **240**, 2646–2656 (2011).
25. Bradley, A. *et al.* The mammalian gene function resource: the International Knockout Mouse Consortium. *Mamm. Genome* **23**, 580–586 (2012).
26. Lee, E. Y. *et al.* Mice deficient for *Rb* are nonviable and show defects in neurogenesis and haematopoiesis. *Nature* **359**, 288–294 (1992).
27. Clarke, A. R. *et al.* Requirement for a functional *Rb-1* gene in murine development. *Nature* **359**, 328–330 (1992).
28. Davis, A. C., Wims, M., Spotts, G. D., Hann, S. R. & Bradley, A. A null *c-myc* mutation causes lethality before 10.5 days of gestation in homozygotes and reduced fertility in heterozygous female mice. *Genes Dev.* **7**, 671–682 (1993).
29. Trumpp, A. *et al.* *c-Myc* regulates mammalian body size by controlling cell number but not cell size. *Nature* **414**, 768–773 (2001).
30. Dubois, N. C. *et al.* Placental rescue reveals a sole requirement for *c-Myc* in embryonic erythroblast survival and hematopoietic stem cell function. *Development* **135**, 2455–2465 (2008).
31. Wu, L. *et al.* Extra-embryonic function of *Rb* is essential for embryonic development and viability. *Nature* **421**, 942–947 (2003).
32. Copp, A. J. Death before birth: clues from gene knockouts and mutations. *Trends Genet.* **11**, 87–93 (1995).
33. Fu, J., Zhao, L., Wang, L. & Zhu, X. Expression of markers of endoplasmic reticulum stress-induced apoptosis in the placenta of women with early and late onset severe pre-eclampsia. *Taiwan. J. Obstet. Gynecol.* **54**, 19–23 (2015).
34. Haider, S. & Knöfler, M. Human tumour necrosis factor: physiological and pathological roles in placenta and endometrium. *Placenta* **30**, 111–123 (2009).
35. Acuna-Hidalgo, R. *et al.* Neu-Laxova syndrome is a heterogeneous metabolic disorder caused by defects in enzymes of the L-serine biosynthesis pathway. *Am. J. Hum. Genet.* **95**, 285–293 (2014).
36. Srivastava, A. *et al.* *De novo* dominant ASXL3 mutations alter H2A deubiquitination and transcription in Bainbridge-Ropers syndrome. *Hum. Mol. Genet.* **25**, 597–608 (2016).
37. Riley, P., Anson-Cartwright, L. & Cross, J. C. The Hand1 bHLH transcription factor is essential for placental and cardiac morphogenesis. *Nat. Genet.* **18**, 271–275 (1998).
38. Adams, R. H. *et al.* Essential role of p38 α MAP kinase in placental but not embryonic cardiovascular development. *Mol. Cell* **6**, 109–116 (2000).
39. Raffel, G. D. *et al.* *Ott1* (*Rbm15*) is essential for placental vascular branching morphogenesis and embryonic development of the heart and spleen. *Mol. Cell. Biol.* **29**, 333–341 (2009).
40. Maruyama, E. O. *et al.* Extraembryonic but not embryonic SUMO-specific protease 2 is required for heart development. *Sci. Rep.* **6**, 20999 (2016).
41. Kozak, K. R., Abbott, B. & Hankinson, O. ARNT-deficient mice and placental differentiation. *Dev. Biol.* **191**, 297–305 (1997).
42. Adelman, D. M., Gertsenstein, M., Nagy, A., Simon, M. C. & Maltepe, E. Placental cell fates are regulated *in vivo* by HIF-mediated hypoxia responses. *Genes Dev.* **14**, 3191–3203 (2000).
43. Linask, K. K., Han, M. & Bravo-Valenzuela, N. J. Changes in vitelline and utero-placental hemodynamics: implications for cardiovascular development. *Front. Physiol.* **5**, 390 (2014).
44. Matthiesen, N. B. *et al.* Congenital heart defects and indices of placental and fetal growth in a nationwide study of 924 422 liveborn infants. *Circulation* **134**, 1546–1556 (2016).

Supplementary Information is available in the online version of the paper.

Acknowledgements We would like to thank N. Karp for expert advice on statistical analyses, I. Sealy for help with PCA analyses, the Flow Cytometry Facility at the Babraham Institute, as well as all contributors to the DMDD programme. This work was supported by Wellcome Trust Strategic Award WT100160MA.

Author Contributions V.P.-G., E.F., A.M., A.S. and M.H. performed the core experiments including histological analyses and TSC work; R.W. performed statistical co-association analyses and DMDD webpage data handling; C.I.M., C.T., J.K.W., E.T., E.J.R., D.G., E.S., H.W.-J. and A.G. performed all mouse colony management, breeding, sample collection and genotyping work; N.S., N.W., J.C. and E.M.B.-N. performed transcriptomics analyses; S.G., W.J.W. and T.M. performed HREM imaging and analyses, J.C.S., E.J.R., D.J.A., T.M. and M.H. designed the study, interpreted results and wrote the manuscript.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Publisher's note: Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations. Correspondence and requests for materials should be addressed to M.H. (myriam.hemberger@babraham.ac.uk).

Reviewer Information *Nature* thanks J. Cross, E. Lacy and the other anonymous reviewer(s) for their contribution to the peer review of this work.

METHODS

Mouse lines. Most mouse lines were generated using the EUCOMM/KOMP knockout first conditional-ready targeted ES cell resource (<http://www.mousephenotype.org/about-ikmc/eucomm-proram/eucomm-targeting-strategies>; targeted trap 'tm1a' allele and null 'tm1b' allele). A few lines were generated by CRISPR-Cas9-mediated gene deletion ('em1' allele). All lines were produced and maintained on a C57BL/6N genetic background at the Wellcome Trust Sanger Institute (<http://www.mousephenotype.org/>) as part of the DMDD project¹⁶. Use of all animals was in accordance with UK Home Office regulations, the UK Animals (Scientific Procedures) Act of 1986 and approved by the Wellcome Trust Sanger Institute's Animal Welfare and Ethical Review Body. Gene knockout lines were designated lethal if no homozygous mutants were present among a minimum of 28 pups at P14 and sub-viable if their proportion fell on or below 13% of total offspring from heterozygous intercrosses⁵. Corresponding cut-off criteria applied to the designation of sub-viability at E14.5. These 'DMDD lines' were assessed at embryonic days E14.5 and/or E9.5, counting the day of the vaginal plug as E0.5. Embryos, placentas and yolk sacs were collected; embryos were processed for HREM imaging¹⁹, placentas were fixed in 4% paraformaldehyde (PFA) and yolk sacs were used for genotyping.

For conditional gene ablation in the embryo proper ('placental rescue'), lines were mated to Flp expressors to generate conditional 'tm1c' alleles (<http://www.mousephenotype.org/about-ikmc/eucomm-program/eucomm-targeting-strategies>), and then crossed with *Sox2-cre* transgenic mice²³. Informative crosses were set up between females carrying at least one conditional allele at the locus of interest and heterozygous males that additionally carried the *Sox2-cre* transgene. Embryos and placentas were collected as before; genotyping was performed on embryonic tail biopsies.

Histology. For histological analysis, at least three mutant and three wild-type placentas from at least two independent litters (with pairs of mutant and wild-type placentas recovered from the same litter if possible) were processed for routine paraffin histology and embedded side-by-side for each strain. No statistical methods were used to predetermine sample size. Placentas of male and female conceptuses were analysed wherever possible. No other randomization is applicable for this study. In all cases, tissue appearance and cellular architecture of the placentas analysed confirmed they were in viable condition even if the associated embryo had been designated as dead or dying. Consecutive 7- μ m sections were produced, and alternate sections mounted. A series of sections per block was processed for haematoxylin and eosin (H&E) staining, using a standard protocol (<https://dmdd.org.uk/placental-analysis-protocols/>). Sections through the sagittal midline were chosen for imaging, indicated at E9.5 by the remnant of the uterine lumen and at E14.5 by the site of insertion of the umbilical cord. Slides were scanned on a Hamamatsu slide scanner and images deposited at <https://dmdd.org.uk>. Phenotypes of placentas were assessed for each strain, blinded for strain viability scores, and recorded by at least two independent investigators. In cases in which all three mutant placentas exhibited a particular abnormality, that defect was scored as a phenotype. In cases in which a defect was unambiguously detected only in two of the initial three placentas analysed, an additional two or three mutant placentas were added to confirm the call. Overall, a phenotype was scored when at least 67% of mutant placentas exhibited that particular abnormality. Criteria for assessing yolk sac morphology encompassed apposition of the visceral yolk sac endoderm and mesoderm layers, and the appearance of blood islands.

Immunostaining and *in situ* hybridization. To gain a more precise view of the structural defects in mutant placentas, mutant placentas from all lines were stained for E-cadherin (Cdh1) to demarcate the labyrinthine syncytiotrophoblast (as well as parietal giant cells at E9.5) and with isolectin BSI-B4 to outline labyrinthine trophoblast and decidua. *In situ* hybridization for *Tpbbp* was used to label the spongiotrophoblast and glycogen cells.

For immunostaining, sections were deparaffinised in xylene and processed through an ethanol series to PBS. Antigen retrieval was performed by boiling in 1 mM EDTA pH 7.2, 0.05% Tween-20 or in 10 mM Na-citrate pH 6.0 buffer followed by blocking in PBS, 0.5% BSA, 0.1% Tween-20. Antibodies used were anti-Cdh1 (1:100; BD Biosciences 610181), anti-laminin (1:100; Sigma L9393), anti-MCT4 (1:100; Merck Millipore AB3314P) and biotin-conjugated isolectin from *Bandeiraea simplicifolia* BSI-B4 (1:100; Sigma L2140). Primary antibodies were detected with appropriate fluorescence or horseradish peroxidase-conjugated secondary antibodies; BSI-B4 was detected with horseradish peroxidase-conjugated streptavidin. Nuclei were counterstained with haematoxylin or DAPI. *In situ* hybridization for *Tpbbp* was performed using a standard protocol¹⁴⁵.

Phenotype data analysis. All genes associated with a placental phenotype in mutant mouse lines were selected for interaction network analysis using esyN (<http://www.esyn.org>), displaying genetic and physical interactions. Expression data for all genes assessed in mouse mutants was obtained by meta-analysis of published RNA sequencing (RNA-seq) datasets²². For test-

ing for co-associations between embryonic and placental defects, two separate analyses were performed. First, we examined the phenotypes of homozygous mutant embryos in which the placentas have been scored for abnormalities (122 embryos), and second, we analysed all homozygous mutant embryos scored for mutant phenotypes (241 embryos) according to placental abnormality observed within the line, which builds on our observation that placental abnormalities were fully penetrant in almost every line. The phenotypes scored in homozygous mutant embryos were summarized into broader phenotype categories within the Mammalian Phenotype Ontology by mapping the phenotype terms recorded onto the DMDD intermediate slim as described⁴⁶.

Statistical analysis used Fisher's exact test to assess for an association or increase in abnormality rate of the phenotypes when placentas were scored as abnormal. An orthogonal potential alpha-star filter was used before the statistical testing to reduce the multiple testing burden as recently described¹⁸. To assess the biological effect of placental abnormalities on the abnormality rate of mutant phenotypes, we followed the previously described procedure¹⁸ of determining the difference in two binomial proportions and calculating the 95% confidence interval using Newcombe's recommended method 10 using the *ci.pd* function of the R Epi package. Significance was adjusted for the effects of multiple testing using the Benjamini-Hochberg procedure to control the false discovery rate at 5%.

Penetrance analysis. The mammalian phenotype (MP) terms assigned during annotation of the embryos were summarized into the ontology slim categories, and the penetrance score of each slim term observed for the line calculated as previously described⁴⁶. The phenotype data was analysed using the DMDD intermediate slim terms, and lower hierarchy slims within ontology terms abnormal brain morphology, abnormal blood vessel morphology, and abnormal heart morphology (Supplementary Table 3).

Generation of mutant TSC lines. The wild-type blastocyst-derived TS-Rs26 TSC line (a gift from the Rossant laboratory, and tested to be mycoplasma-free) was cultured as described previously^{20,21}. Differentiation was induced by culturing in media lacking bFGF, heparin and embryonic fibroblast-conditioned medium.

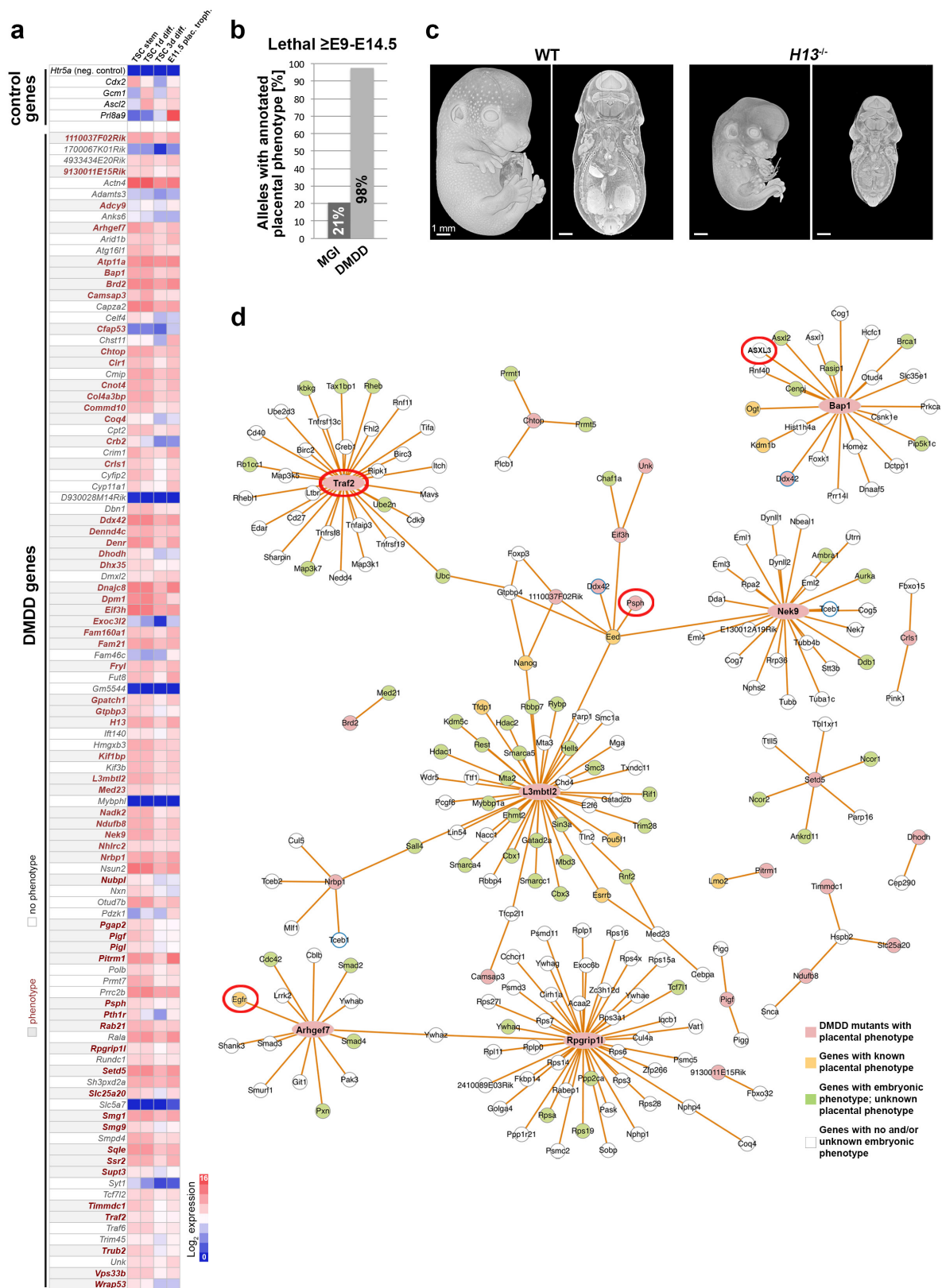
For generation of CRISPR-Cas9-mediated knockout TSCs, guide RNAs (gRNAs) that result in frameshift mutations were designed using the CRISPR design tool (<http://CRISPR.mit.edu>) and checked for high specificity by nucleotide blast searches. gRNA sequences were cloned into the Cas9.2A-eGFP plasmid (Plasmid 48138 Addgene) and verified by sequencing. Empty vector Cas9.2A-eGFP and gene-specific gRNA plus Cas9.2A-eGFP constructs were used to generate vector control and knockout TSCs (Extended Data Fig. 6). Transfection was carried out with Lipofectamine 2000 (ThermoFisher Scientific 11668019) reagent according to the manufacturer's protocol. Knockout clones were confirmed by genotyping using primers spanning the deleted exon, and by reverse transcription followed by semi-quantitative PCR (RT-qPCR) with primers within, and downstream of, the deleted exon, as shown (Extended Data Fig. 6). Five or six independent knockout clones were analysed for each gene mutation.

RT-qPCR expression analysis. Potential defects in TSC maintenance and differentiation capacity were investigated by analysing the expression levels and dynamics of trophoblast marker genes in mutant and control TSCs in stem-cell conditions and following 3 and 6 days of differentiation. Total RNA was extracted using TRI reagent (Sigma T9424), DNase-treated and 1 μ g used for cDNA synthesis with RevertAid H-Minus reverse transcriptase (Thermo Scientific EP0451). Quantitative PCR was performed using SYBR Green Jump Start Taq Ready Mix (Sigma S4438) and intron-spanning primer pairs²¹ (Supplementary Table 4) on a Bio-Rad CFX96 or CFX384 thermocycler. Normalized expression levels are displayed as mean relative to the vector control sample; error bars indicate s.e.m. of at least three replicates.

Transcriptomics analysis. Samples were lysed in TRI reagent with a 5-mm stainless steel bead (Qiagen) for 4 min at 20 Hz in a tissue lyser (Qiagen). After chloroform extraction for 30 min at room temperature, RNA was extracted from the aqueous phase using ethanol and a spin column (Qiagen RNeasy MinElute). After quantification (Qubit RNA BR) the sample was treated with DNase enzyme (Qiagen) and purified over a spin column. Adaptor indexed strand-specific RNA-seq libraries were generated from 1 μ g of total RNA following the dUTP method using the stranded mRNA LT sample kit (Illumina). Libraries were pooled and sequenced on Illumina HiSeq 2000 in 75-base-pair paired-end mode. FASTQ files were aligned to the GRCh38.p5 reference genome using TopHat (v2.0.13, options: -library-type fr-firststrand). Counts for genes were produced using htseq-count (v0.6.1 options: -stranded = reverse) with the Ensembl v90 annotation as a reference. The data were assessed for technical quality (GC content, insert size and gene body coverage) using QoRTs⁴⁷ and poor quality samples were removed. A variance-stabilizing transformation was applied to count data for each gene using the R package DESeq2 varianceStabilizingTransformation function⁴⁸. Principal component analysis was performed on the transformed count data for each gene using the R *prcomp* function.

Data availability. All placental phenotyping data are available at <https://dmdd.org.uk>. Sequence data were deposited in the European Nucleotide Archive (ENA) under accession ERP023265. All primers sequences are provided. All other data are available from the corresponding author upon reasonable request.

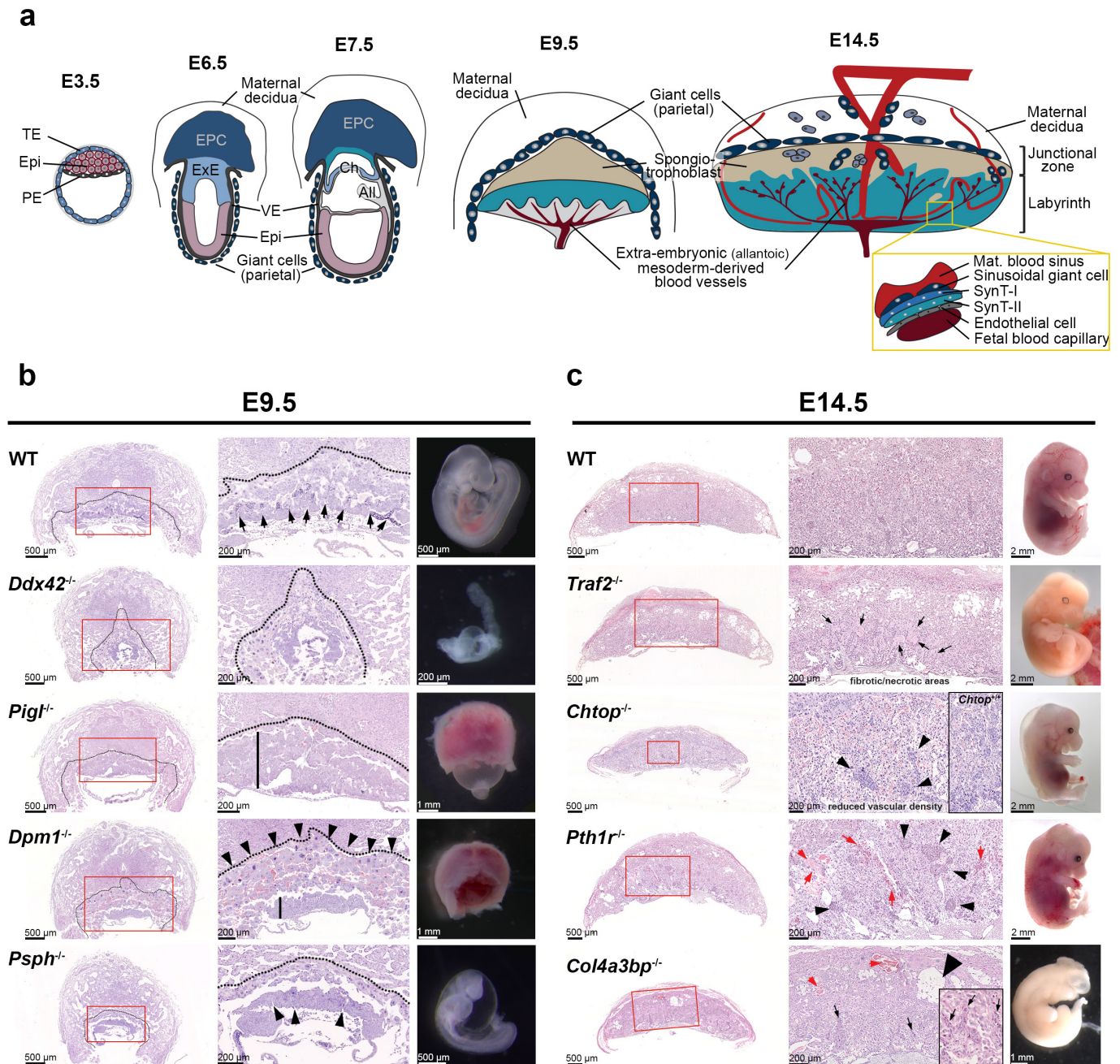
45. Hemberger, M., Nozaki, T., Masutani, M. & Cross, J. C. Differential expression of angiogenic and vasodilatory factors by invasive trophoblast giant cells depending on depth of invasion. *Dev. Dyn.* **227**, 185–191 (2003).
46. Wilson, R., McGuire, C., Mohun, T. & Project, D. Deciphering the mechanisms of developmental disorders: phenotype analysis of embryos from mutant mouse lines. *Nucleic Acids Res.* **44** (D1), D855–D861 (2016).
47. Hartley, S. W. & Mullikin, J. C. QoRTs: a comprehensive toolset for quality control and data processing of RNA-seq experiments. *BMC Bioinformatics* **16**, 224 (2015).
48. Love, M. I., Huber, W. & Anders, S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol.* **15**, 550 (2014).



Extended Data Figure 1 | See next page for caption.

Extended Data Figure 1 | Potential trophoblast gene function in mutants with placental defect. **a**, Expression of trophoblast control genes and the 103 DMDD genes in TSCs, TSCs differentiated for 1 or 3 days (d), and in E11.5 placentas. \log_2 -transformed expression values (read counts per million) of RNA-seq data are displayed. Note that all genes associated with a placental phenotype in mutants (labelled in red font) are expressed in trophoblast. **b**, Frequency of placental defects annotated in mid-gestational lethal mutants (MP: 0011098) as annotated in the MGI database, compared to the findings in DMDD in which 40 out of 41 E9.5–E14.5 embryonic lethal mice were found to exhibit placental abnormalities. **c**, Left, volume rendered 3D model of the surface of a

wild-type embryo, staged as Theiler stage (TS) 23, and coronal section through the volume rendered model. Right, equivalent images of a littermate E14.5 *H13*^{-/-} embryo, staged as TS21. Note that the models are displayed in identical resolutions. Scale bar, 1 mm. Images are representative of at least five embryos per genotype. **d**, Network analysis using esyN (<http://www.esyn.org>) for all DMDD genes identified as causing a placental phenotype in mutants. BAP1 and ASXL3 are known interactors in humans. Red circles identify genes implicated in human trophoblast-based pathologies. The analysis reveals molecular nodes that appear to be of key importance for placental development.

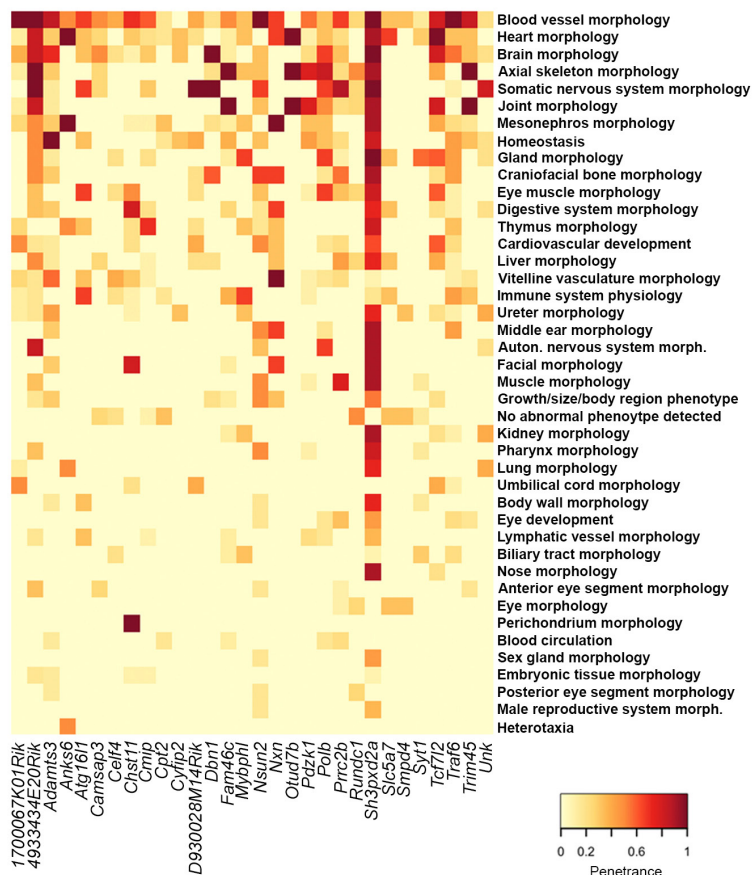


Extended Data Figure 2 | Identification of placental defects by haematoxylin and eosin histology. **a**, Schematic representation of key stages and cell types in extraembryonic development, complementing Fig. 2c, d. All, allantois; Ch, chorion; Epi, epiblast; ExE, extraembryonic ectoderm; PE, primitive endoderm; SynT-I and SynT-II, syncytiotrophoblast layers I and II; TE, trophoblast; VE, visceral endoderm. **b**, Examples of E9.5 placental phenotypes. Dotted lines denote the boundary to maternal decidua; vertical bars denote chorion trophoblast thickness; arrows in the wild-type placenta indicate invagination sites of extraembryonic mesoderm-derived blood vessels

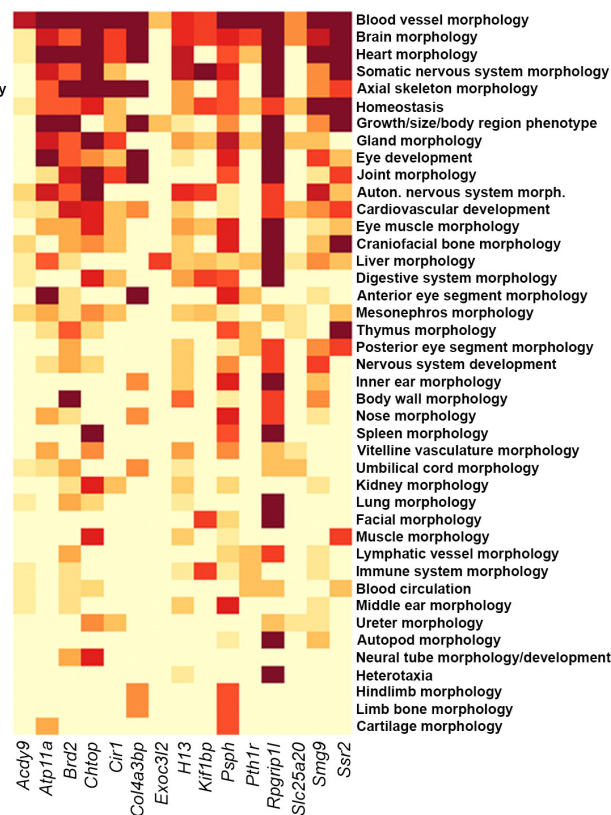
into chorionic trophoblast; arrowheads in the *Psph*^{-/-} placenta denote sites of chorion folding but missing blood vessels; arrowheads in the *Dpm1*^{-/-} placenta denote overabundant and enlarged trophoblast giant cells. **c**, Examples of E14.5 placental phenotypes. Red arrows indicate abnormal maternal blood accumulations. Arrows in *Traf2*^{-/-} and *Col4a3bp*^{-/-} (including inset) placentas denote fibrotic and/or necrotic areas; arrowheads in *Chtop*^{-/-} and *Pth1r*^{-/-} placentas indicate abnormal spongiotrophoblast inclusions. Representative mutant embryo images are also depicted. Images of mutant placentas in **b** and **c** are representative of at least three independent mutants per line, see Methods.

a

Normal Placenta



Abnormal Placenta

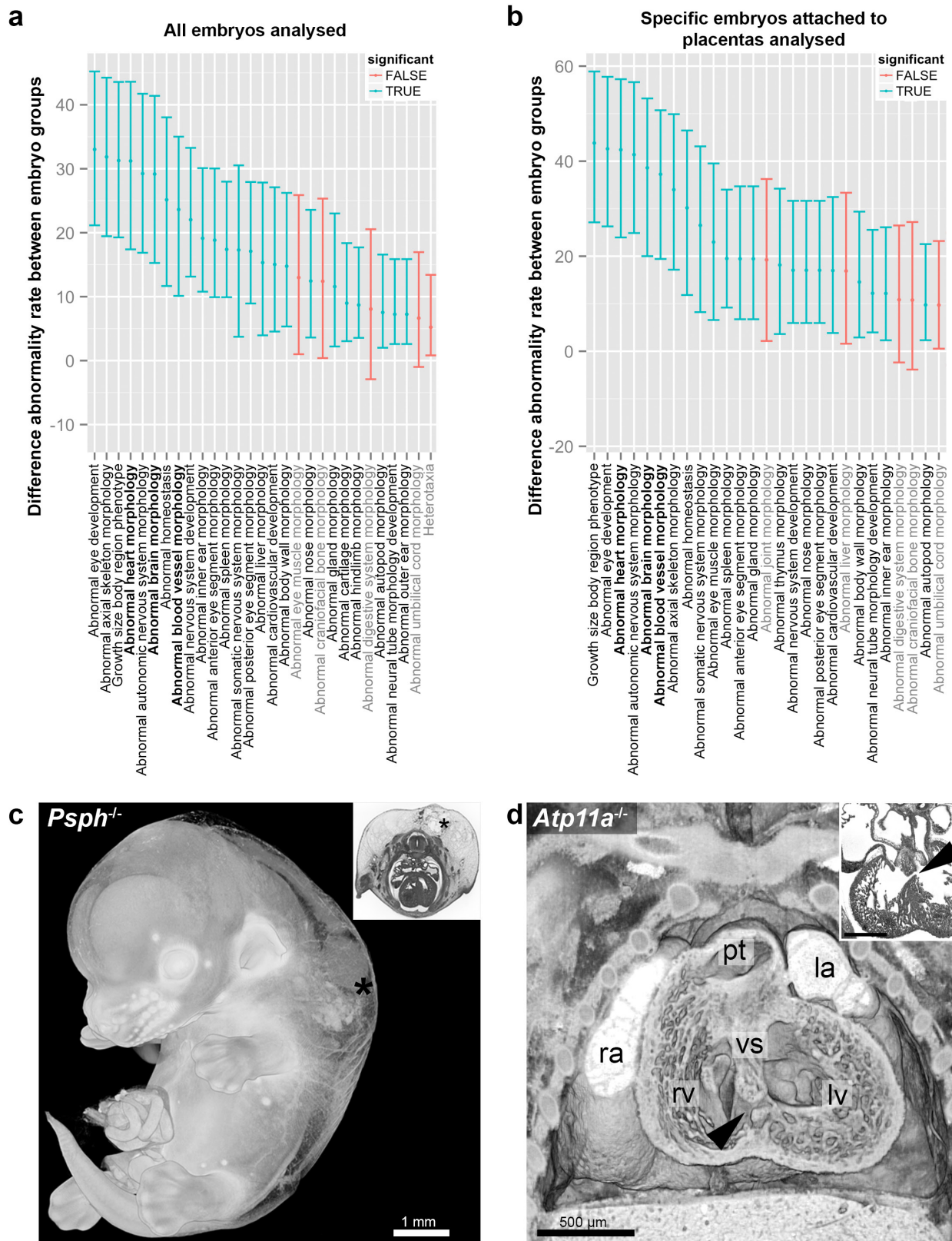


b

| Embryonic defects enriched in lines with an abnormal placenta | % difference | p value |
|---------------------------------------------------------------|--------------|----------|
| Growth/size phenotype | 31.28 | 1.93E-06 |
| Heart morphology | 31.22 | 5.18E-05 |
| Brain morphology | 29.17 | 2.49E-04 |
| Blood vessel morphology | 23.62 | 3.17E-03 |
| Subcutaneous edema | 22.25 | 8.64E-03 |
| Nervous system development | 22.03 | 1.26E-06 |

Extended Data Figure 3 | Co-association analysis between embryo and placenta phenotypes. a, Mutant mouse lines were classified into those that exhibit a placental phenotype at E14.5 and those that do not. All embryos analysed by HREM imaging were tagged accordingly to either of these two groups. Enrichment of embryonic phenotype terms in mutant strains with normal or abnormal placentas is shown (dark red denotes fully penetrant phenotype). For brevity, the 'abnormal' description has been removed from ontology terms. b, Significantly enriched embryonic

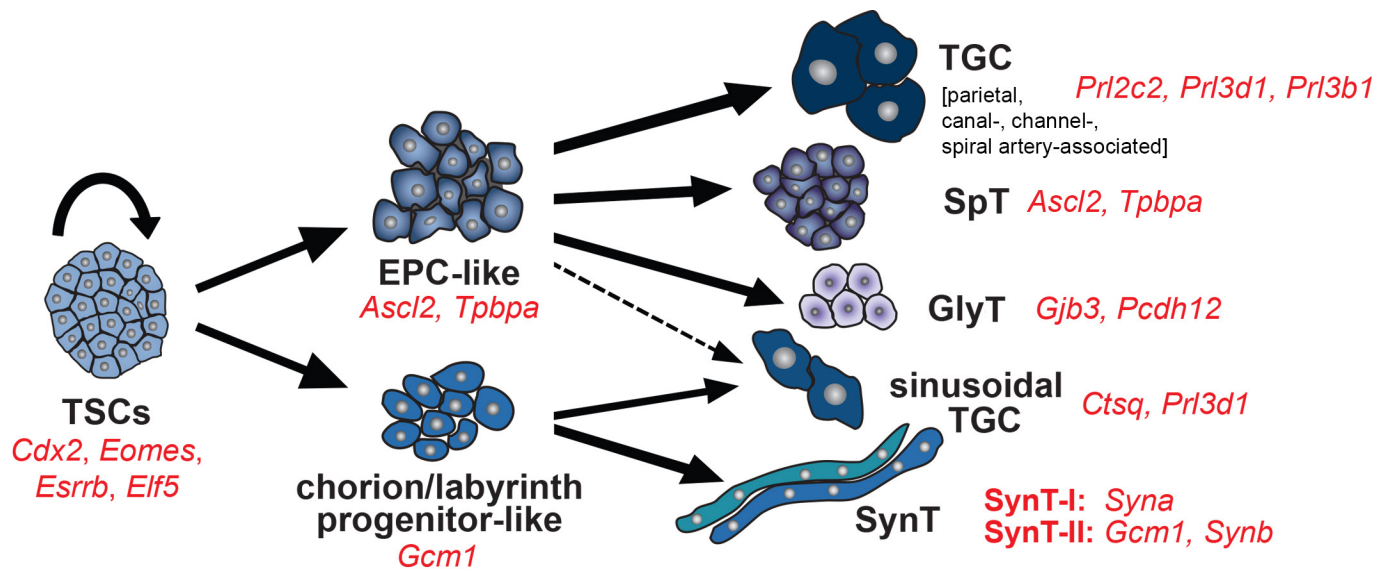
phenotype terms in lines that exhibit an abnormal placenta (see also Supplementary Table 2) versus those with normal placenta. Following hypothesis testing using Fisher's exact test, adjusting for multiple testing using the Benjamini–Hochberg method, we estimated the magnitude of the abnormal placenta effect. This was determined by calculating independent binomial proportions for the two groups of embryos with normal ($n = 172$) and abnormal ($n = 69$) placenta. The percentage difference between groups and the P values are shown.



Extended Data Figure 4 | See next page for caption.

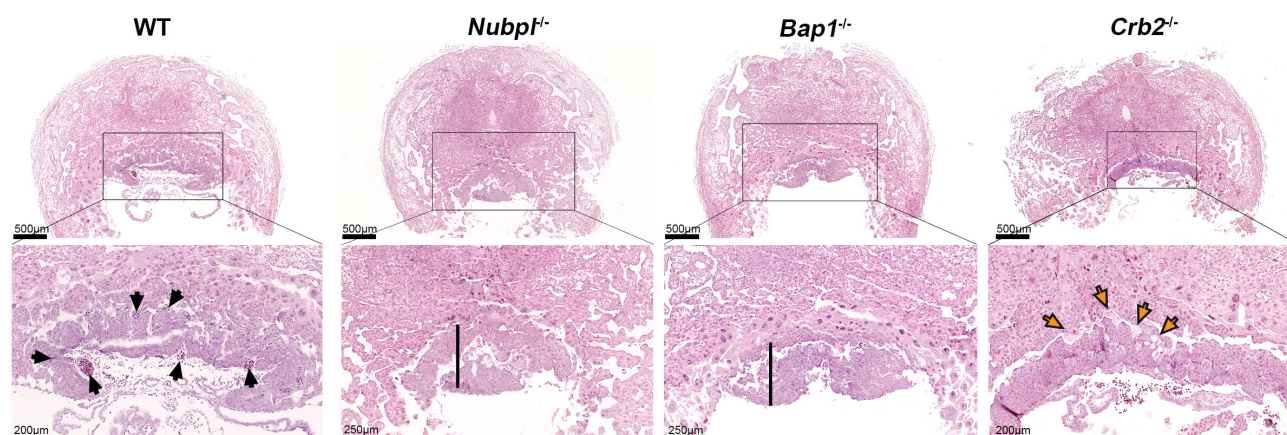
Extended Data Figure 4 | Specific embryonic defects are significantly correlated with the occurrence of an abnormal placenta. **a**, Further, detailed co-association statistics between the occurrence of a placental phenotype and specific abnormalities in the embryo proper in DMDD lines. As before, mutant mouse lines were classified into those that exhibit a placental phenotype at E14.5 and those that do not. All embryos analysed by HREM imaging were tagged accordingly to either of these two groups. Significant differences in the frequency of specific embryonic defects was determined between these two groups, and scored for the size of the effect and for its significance. Following hypothesis testing using Fisher's exact test, adjusting for multiple testing using the Benjamini–Hochberg method, we estimated the magnitude of the abnormal placenta effect. This was determined by calculating independent binomial proportions for the two groups of embryos with normal ($n = 172$) and abnormal ($n = 69$) placenta. The figure shows the differences in the estimated abnormality rates of the two embryo groups, and the extent of the bars represent the 95% Newcombe confidence interval (see Methods). 'True' means that these associations are significant, 'false' that they fall below the significance

threshold. Please note that some terms, such as eye development and growth/size/body region, are probably a consequence of developmental retardation. However, the highlighted terms such as heart, brain and vascular system morphology are definitely based on abnormalities that are not merely due to developmental delay. **b**, Same analysis as in **a** but only including those specific embryos whose placentas that were analysed histologically (as opposed to all embryos per strain; $n = 81$ and $n = 41$ embryos having normal and abnormal placenta, respectively). Note that the important and meaningful terms hold up to significance irrespectively. **c**, HREM image of an example of a massive subcutaneous oedema (asterisk) covering the entire back of a *Psph*^{-/-} embryo. Volume rendered 3D model. Axial section through the level of the heart is shown as inlay. Note also the delay in developmental progress. **d**, Muscular ventricular septal defect (arrowhead) in an *Atp11a*^{-/-} embryo. Coronal section through volume rendered 3D model. Axial HREM image is shown as inlay. la, left atrial appendix; lv, left ventricle; pt, pulmonary trunk; ra, right atrial appendix; rv, right ventricle; vs, ventricular septum. Embryo defects shown in **c** and **d** are representative of at least three independent mutants.

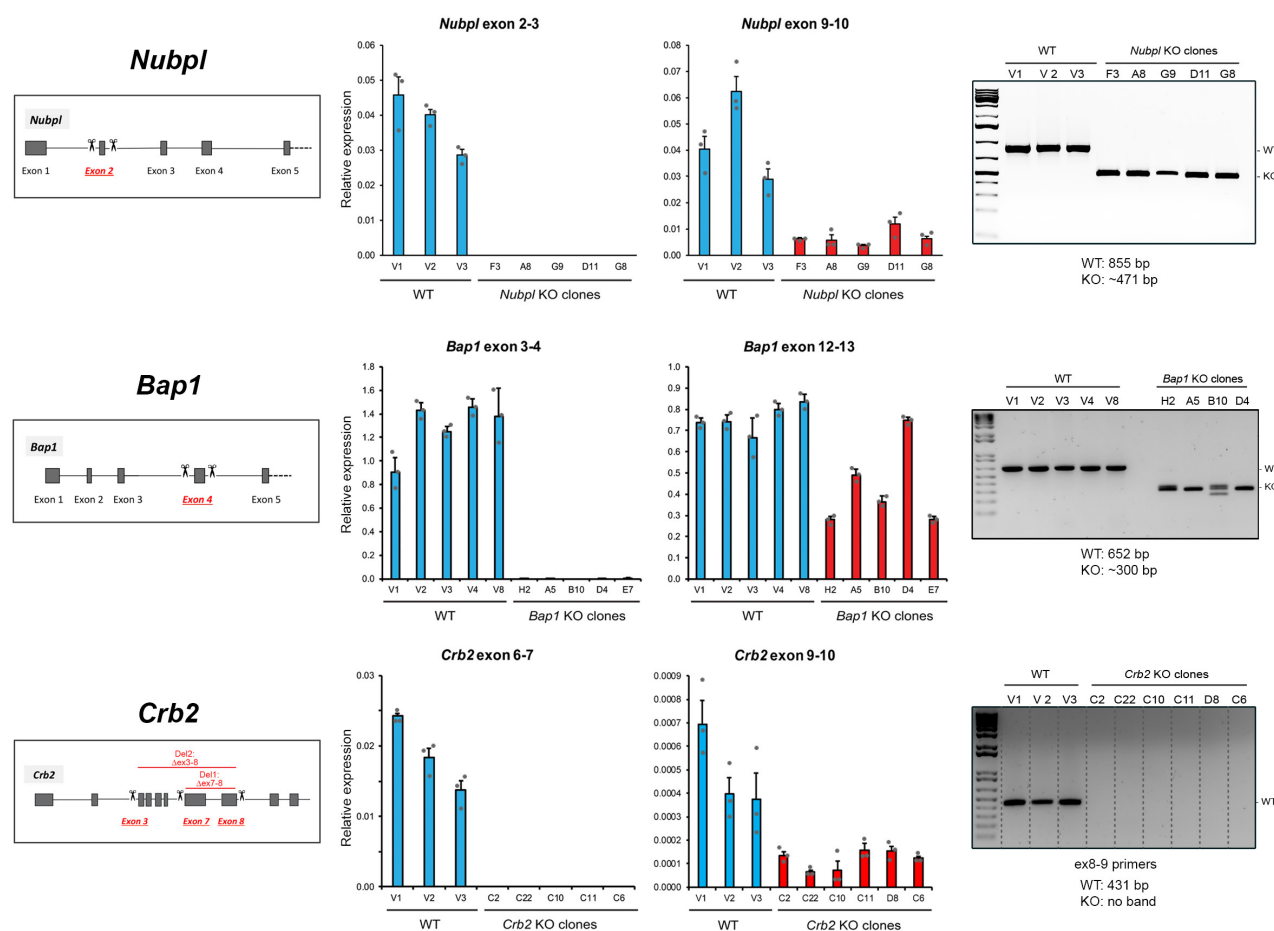


Extended Data Figure 5 | Major routes of TSC differentiation. Diagram of the main differentiation routes of TSCs, including representative cell type-specific marker genes.

a

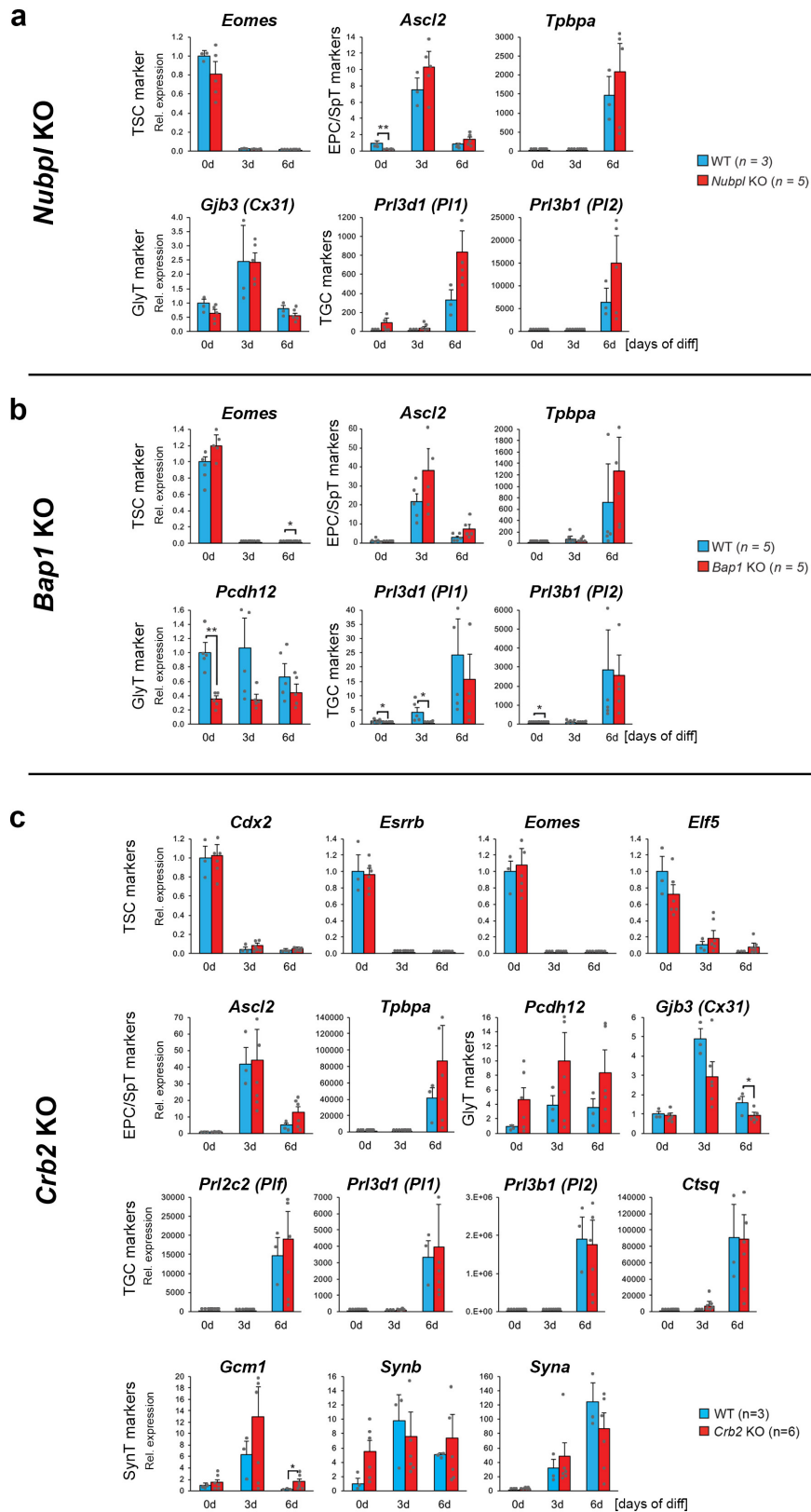


b



Extended Data Figure 6 | Selection of genes for in-depth analysis of trophoblast contribution to embryonic lethality. a, E9.5 phenotypes of mutant placentas of the three genes (*Nubpl*, *Bap1* and *Crb2*) chosen for ablation in TSCs, as well as for placental rescue analysis *in vivo* (Fig. 5, Extended Data Figs 8–10). Black arrows (wild-type placenta) denote fetal blood vessels penetrating into the chorionic ectoderm. Vertical bars denote unpatterned appearance of chorion. Orange arrows indicate empty or fibrotic maternal blood spaces. Images are representative of at least three mutants per line. **b**, Details of CRISPR design and TSC clone

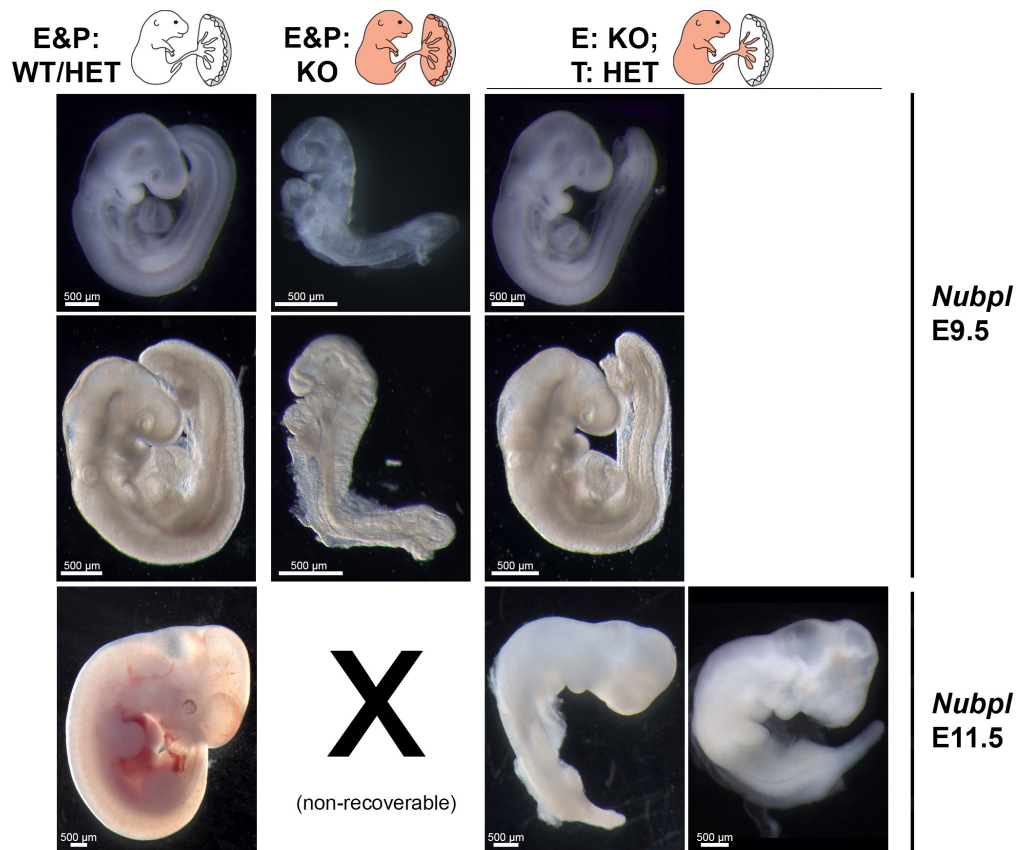
screening strategy for the three selected genes *Nubpl*, *Bap1* and *Crb2*. All targeted exons were first confirmed to be expressed in trophoblast. RT-qPCR (performed in technical triplicate per clone) and genomic genotyping PCR analysis (performed in duplicate per sample, with results independently confirmed by RT-qPCR data) were performed on individual, single-cell expanded TSC clones to confirm homozygous knockout. Of note, even though splicing may occur across the deleted exon, all CRISPR-Cas9 deletions were designed to result in a premature stop codon. RT-qPCR data are mean \pm s.e.m. of $n = 3$ technical replicates.



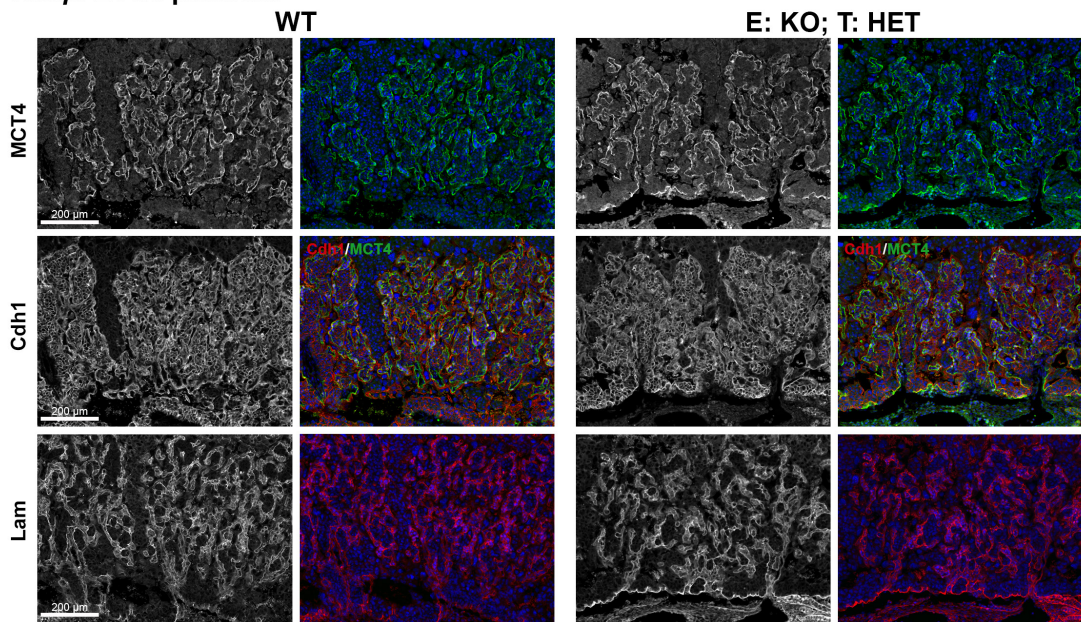
Extended Data Figure 7 | Analysis of mutant TSCs for defects in TSC maintenance and differentiation. a, *Nubpl*^{-/-} TSC clones assessed for additional trophoblast marker genes by RT-qPCR. b, Additional marker gene analysis on *Bap1*-mutant TSCs. c, Analysis of *Crb2*^{-/-} TSC clones for a phenotype in stem-cell maintenance (0 days) or during differentiation

(3 or 6 days). No significant difference in cell morphology, growth behaviour and gene expression pattern was observed compared to wild-type vector control clones. Data are mean \pm s.e.m. * $P < 0.05$; ** $P < 0.01$ (ANOVA with Holm-Bonferroni's post-hoc test).

a

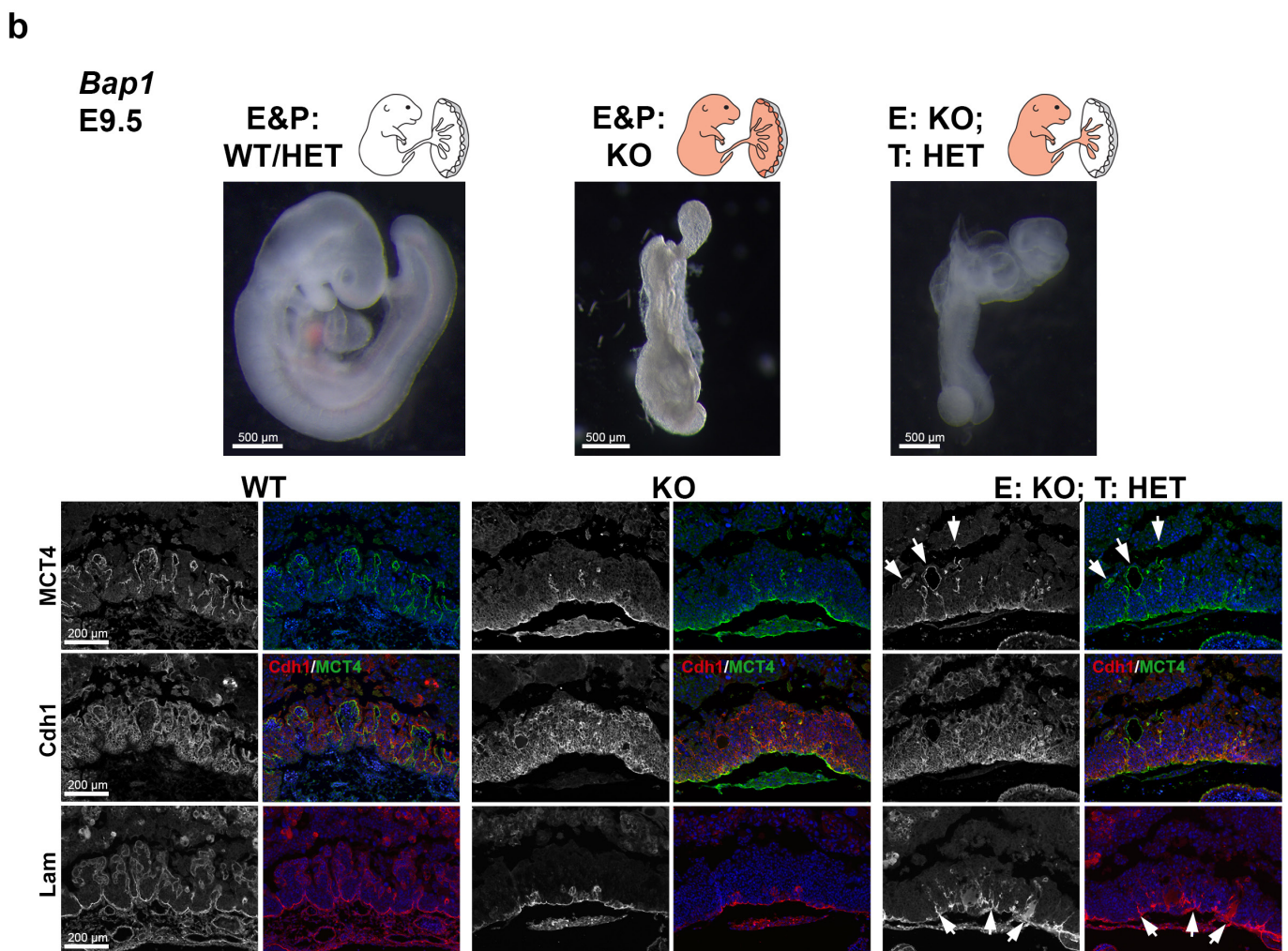
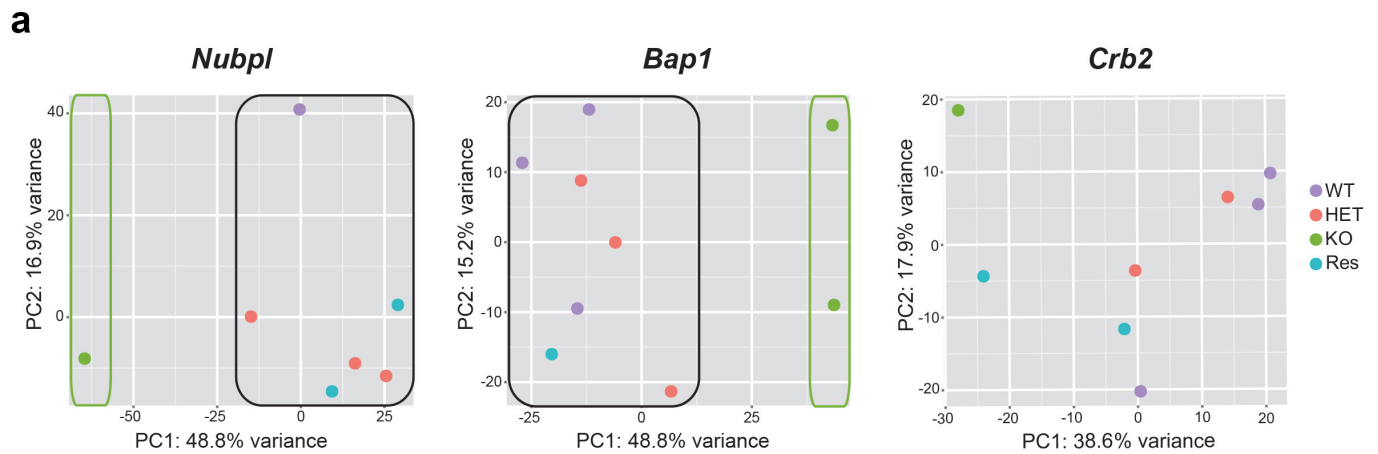


b

***Nubpl* E11.5 placenta**

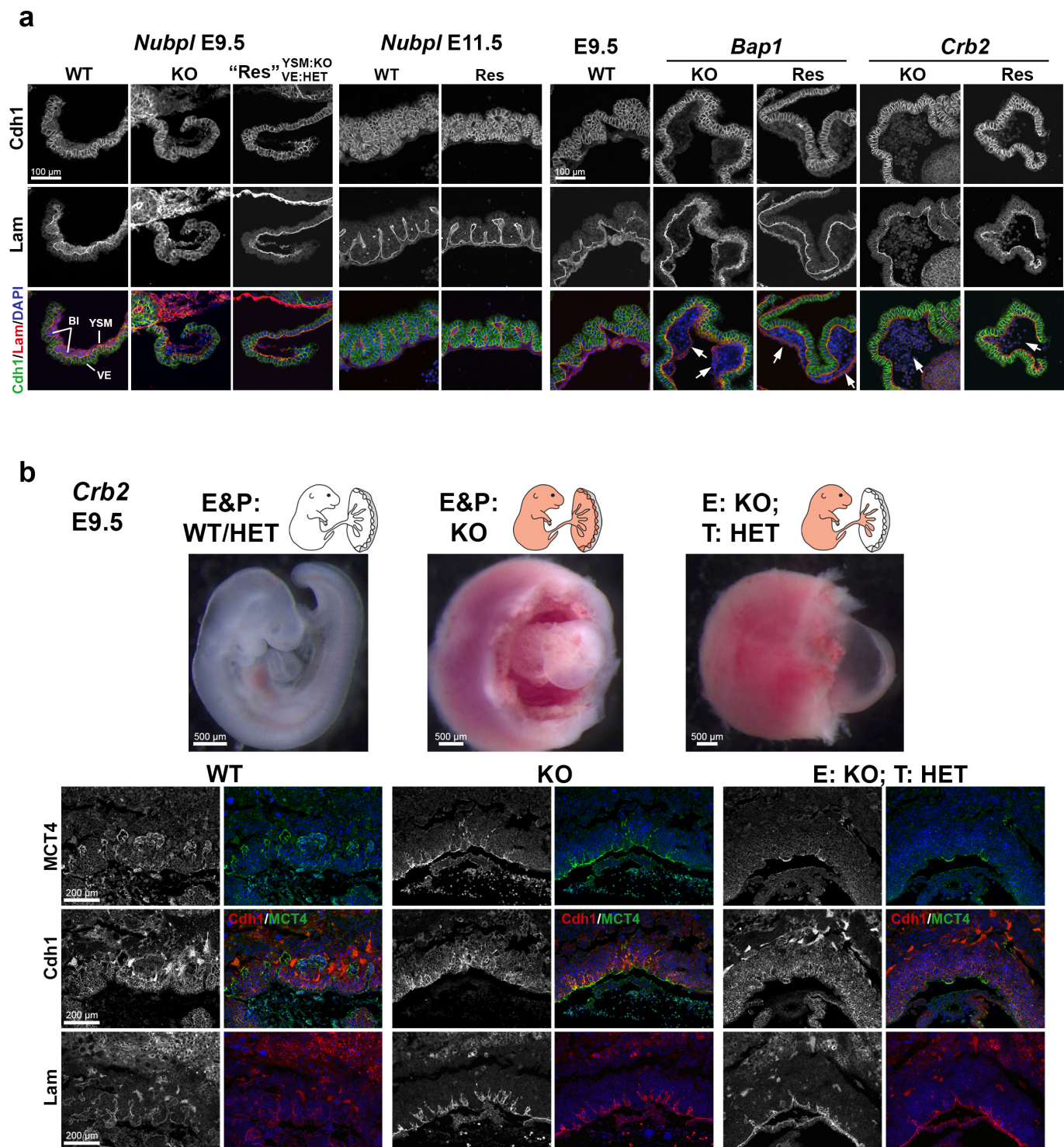
Extended Data Figure 8 | Placental rescue of *Sox2-cre*-mediated conditional knockout of *Nubpl*. **a**, Additional images of *Nubpl*-mutant embryos showing that a wild-type trophoblast compartment rescues the developmental retardation phenotype and embryonic defects observed in the full knockout at E9.5. At E11.5, *Nubpl*^{-/-} embryos can still be recovered, whereas complete knockout embryos are not retrievable any more. Images are representative of at least ten independent embryos with

the corresponding genotype. **b**, Histological analysis of the corresponding placentas at E11.5 shows a complete rescue of the placental defect in conditional knockouts with a genetically functional trophoblast lineage. Sections were stained for MCT4 (a SynT-II marker), E-cadherin (a global SynT marker) and laminin (blood vessel basement membrane marker). Images are representative of three placentas per genotype.



Extended Data Figure 9 | Transcriptomic analysis of placentas from rescue experiments and developmental performance of *Bap1* conditional knockouts. **a**, Principal component analysis of global transcriptomes of E9.5 placentas with the indicated genotype. 'Res' refers to placentas from *Sox2-cre*-mediated conditional knockouts in which the trophoblast lineage remains functional, whereas the embryo is ablated for the gene of interest (E: KO; T: HET). **b**, Top, E9.5 embryo photos of the depicted genotypes for the *Bap1* strain. The embryonic lethality of the

complete *Bap1* knockout cannot be rescued by a functional trophoblast compartment. Images are representative of at least 12 independent embryos per genotype. Bottom, histological analysis of the corresponding placentas, stained as in Fig. 5b and Extended Data Fig. 8b. Arrows point to partially rescued syncytiotrophoblast loops and some vascular invaginations into the chorionic ectoderm. Yet the vascularisation of the forming labyrinth layer remains under-developed compared to controls. Images are representative of three placentas per genotype.



Extended Data Figure 10 | Analysis of yolk sac morphology in *Nubpl*, *Bap1* and *Crb2* mutants and developmental performance of *Crb2* conditional knockout. **a, Immunofluorescence staining of yolk sacs for E-cadherin (green) and laminin (red) demarcating the visceral endoderm (VE) and basement membrane of the yolk sac mesoderm (YSM), respectively. BI, blood cells. *Bap1* and *Crb2* mutants show a defect characterized by the lack of attachment of the two visceral yolk sac layers (arrows). This defect cannot be rescued by the *Sox2-cre*-mediated**

conditional knockout, indicating that its cause resides in the extra-embryonic mesoderm lineage. **b**, Developmental performance of *Crb2* knockout and conditional knockout embryos and analysis of placental morphology, equivalent to Extended Data Figs 8b and 9b. No rescue of embryonic lethality or placental defects is observed in the conditional knockouts. Images are representative of at least three independent conceptuses per genotype.

DNA methylation-based classification of central nervous system tumours

A list of authors and their affiliations appears in the online version of the paper.

Accurate pathological diagnosis is crucial for optimal management of patients with cancer. For the approximately 100 known tumour types of the central nervous system, standardization of the diagnostic process has been shown to be particularly challenging—with substantial inter-observer variability in the histopathological diagnosis of many tumour types. Here we present a comprehensive approach for the DNA methylation-based classification of central nervous system tumours across all entities and age groups, and demonstrate its application in a routine diagnostic setting. We show that the availability of this method may have a substantial impact on diagnostic precision compared to standard methods, resulting in a change of diagnosis in up to 12% of prospective cases. For broader accessibility, we have designed a free online classifier tool, the use of which does not require any additional onsite data processing. Our results provide a blueprint for the generation of machine-learning-based tumour classifiers across other cancer entities, with the potential to fundamentally transform tumour pathology.

The developmental complexity of the brain is reflected by the vast array of distinct brain tumour entities defined in the current WHO (World Health Organization) classification of central nervous system (CNS) tumours¹. These tumours are clinically and biologically highly diverse, encompassing a wide spectrum from benign neoplasms, which can frequently be cured by surgery alone (for example, pilocytic astrocytoma), to highly malignant tumours that respond poorly to any therapy (for example, glioblastoma). Previous studies have reported substantial inter-observer variability in the histopathological diagnosis of many CNS tumours, for example, in diffuse gliomas², ependymomas³ and supratentorial primitive neuroectodermal tumours⁴. To address this, some molecular grouping has been introduced into the update of the WHO classification, but only for selected entities such as medulloblastoma. Furthermore, several single-gene tests based on DNA methylation analysis (for example, *MGMT* promoter methylation status), fluorescence *in situ* hybridization (for example, 1p/19q codeletion, *EGFR*, *MYC*, *MYCN*, *PDGFRA*, 19q13.42) or immunohistochemistry (for example, CTNNB1 and LIN28A) that are required to cover the most important differential diagnoses have been shown to be difficult to standardize. Such diagnostic discordance and uncertainty may confound decision-making in clinical practice as well as the interpretation and validity of clinical trial results.

The cancer methylome is a combination of both somatically acquired DNA methylation changes and characteristics that reflect the cell of origin^{5,6}. The latter property enables, for example, the tracing of the primary site of highly dedifferentiated metastases of cancers of unknown origin⁷. It has been convincingly shown that DNA methylation profiling is highly robust and reproducible even from small samples and poor quality material⁸, and such profiles have been widely used to subclassify CNS tumours that were previously considered homogeneous diseases^{4,9–16}. On the basis of this previous work within single entities, we present a comprehensive approach for the DNA methylation-based classification of all CNS tumour entities across age groups.

CNS tumour reference cohort

To establish a comprehensive CNS tumour reference cohort, we generated genome-wide DNA methylation profiles using Infinium HumanMethylation450K BeadChip arrays (minimum of eight cases

per group) representing almost all WHO-defined neuroectodermal and sellar region tumours¹. We further profiled mesenchymal tumours, melanoma, diffuse large B-cell lymphoma, plasmacytoma and six types of pituitary adenomas, in total comprising 76 histopathological entities and seven entity variants that occur in the CNS. All histopathological entities and variants were analysed by unsupervised clustering both within each entity and across histologically similar tumour entities, aiming to identify (i) distinct DNA methylation classes within one histopathological entity and (ii) DNA methylation classes comprising tumours displaying a varied histological phenotype. This iterative process led to the designation of 82 CNS tumour classes characterized by distinct DNA methylation profiles (Fig. 1a). Of these, 29 classes were equivalent to a single WHO entity (category 1); 29 classes represented subclasses within a WHO entity (category 2); in 8 classes, WHO grading was not fully recapitulated (category 3) and in 11 classes, the boundaries of methylation classes were not identical to the entity boundaries of WHO (category 4) (Fig. 1a). The remaining five represented DNA methylation classes that have not been defined by the WHO classification (category 5), three of which were recently described⁴ as well as the not yet well-defined class of astrocytoma and one new subclass of infantile hemispheric glioma. There was evidence for several additional classes of rare tumours, with too few cases to be included at present. Taking the impact of the tumour microenvironment on the methylation profile into consideration, we included 47 tumour samples with a pronounced inflammatory or reactive tumour microenvironment, both of which have distinct methylation profiles. We additionally selected 72 samples that represent seven non-neoplastic CNS regions, resulting in a combined reference cohort of 2,801 samples from 91 classes (Fig. 1a) that was visualized using *t*-distributed stochastic neighbour embedding (*t*-SNE) dimensionality reduction¹⁷ (Fig. 1b). This analysis further supported the separation of samples into the defined DNA methylation classes (see also Extended Data Fig. 1a, b; unprocessed IDAT files can be downloaded from the NCBI Gene Expression Omnibus (GEO), under accession number GSE109381). Supplementary Table 1 gives an overview of methylation class characteristics and Supplementary Table 2 shows case-by-case information of the reference samples.

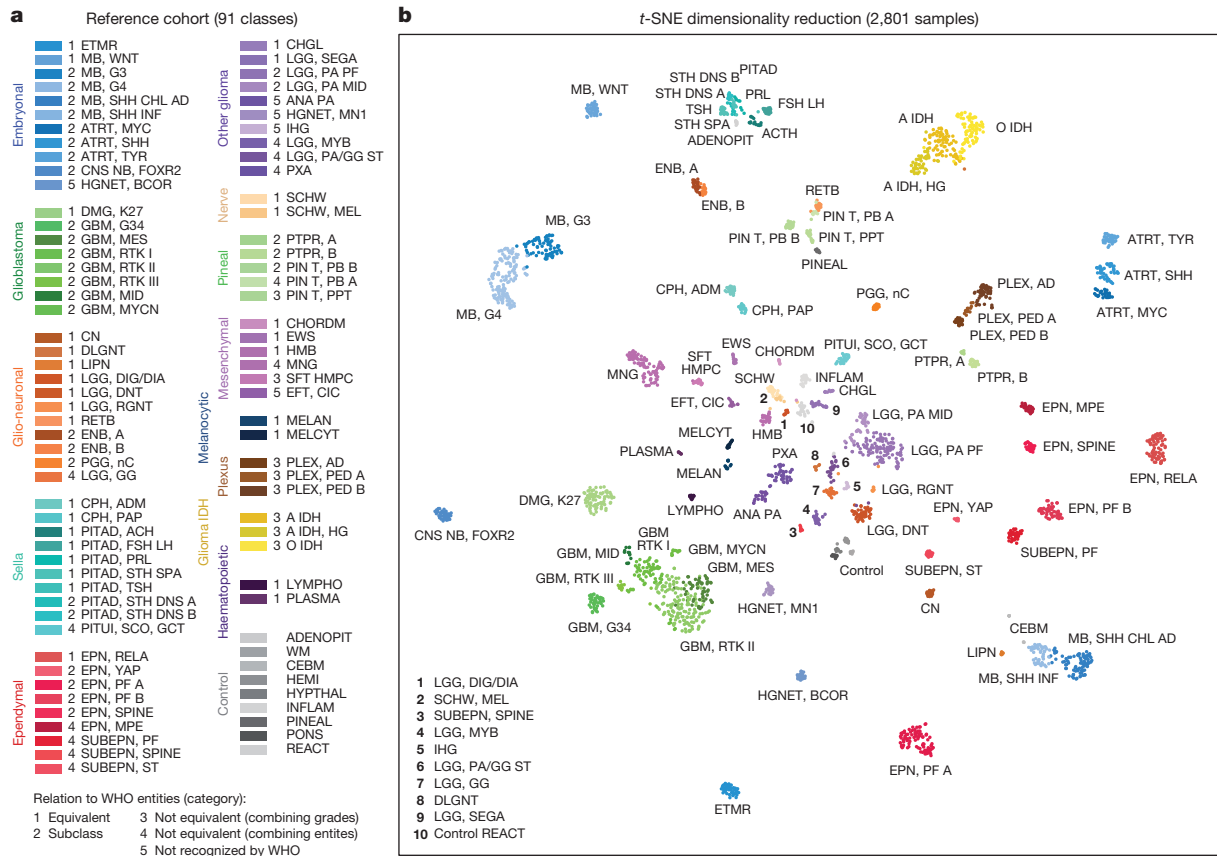


Figure 1 | Establishing the DNA methylation-based CNS tumour reference cohort. **a**, Overview of the 82 CNS tumour methylation classes and nine control tissue methylation classes of the reference cohort. The methylation classes are grouped by histology and colour-coded. Category 1 methylation classes are equivalent to a WHO entity, category 2 methylation classes are a subgroup of a WHO entity, category 3 methylation classes are not equivalent to a unique WHO entity with combining of WHO grades, category 4 methylation classes are not equivalent to a unique WHO entity with combining of WHO entities, and category 5 methylation classes are not recognized as a WHO entity. Full names and further details of the 91 classes are included in Supplementary Table 1. Embryonal tumours, shades of blue; glioblastomas, shades of

green; other gliomas, shades of violet; ependymomas, shades of red; glio-neuronal tumours, shades of orange; IDH-mutated gliomas, shades of yellow; choroid plexus tumours, shades of brown; pineal region tumours, shades of mint green; melanocytic tumours, shades of dark blue; sellar region tumours, shades of cyan; mesenchymal tumours, shades of pink; nerve tumours, shades of beige; haematopoietic tumours, shades of dark purple; control tissues: shades of grey. **b**, Unsupervised clustering of reference cohort samples ($n = 2,801$) using *t*-SNE dimensionality reduction. Individual samples are colour-coded in the respective class colour ($n = 91$) and labelled with the class abbreviation. The colour code and abbreviations are identical to **a**.

Classifier development

Application in routine diagnostics requires fast and reproducible classification of samples as well as a measure of confidence for the specific call. To this end, we used the random forest algorithm, which is a so-called ensemble method that combines the predictions of several 'weak' classifiers to achieve improved prediction accuracy¹⁹. Using this algorithm, we generated 10,000 binary decision trees, incorporating genome-wide information from all 2,801 reference samples and 91 methylation classes (Extended Data Fig. 4). Each of these trees assigns a given diagnostic sample to one of the 91 classes, resulting in an aggregate raw score (Fig. 2a). To obtain class probability estimates that can

be used to guide diagnostic decision-making, we fitted a multinomial logistic regression calibration model that transforms the raw score into a probability that measures the confidence in the class assignment (the calibrated score). The calibration enables the comparison of classifier results between classes despite a different raw score distribution (Extended Data Fig. 5a, b). Cross-validation of the random forest classifier resulted in an estimated error rate of 4.89% for raw scores and 4.28% for calibrated scores and an area under receiver operating characteristic curve of 0.99, indicating a high discriminating power (Fig. 2b and Extended Data Fig. 5c). The vast majority of cross-validation misclassifications occurred within eight groups of histologically and biologically closely related tumour classes, the distinction of which is currently without clinical impact (with the possible exception of choroid plexus tumours¹³; Fig. 2b). We therefore defined eight 'methylation class families', for which calibrated scores are summed up to a single score. This reduced the cross-validated error rate for the clinically relevant groupings to 1.14% (Fig. 2b and Extended Data Fig. 5c). Taking the maximum score for class assignment and using a multiclass approach²⁰, overall sensitivity and specificity was 0.989 and 0.999, respectively (Extended Data Fig. 5c).

For application to diagnostic tumour samples, a threshold value for the prediction of a matching class is required. Using receiver operating characteristic curve analysis of the maximum calibrated scores we

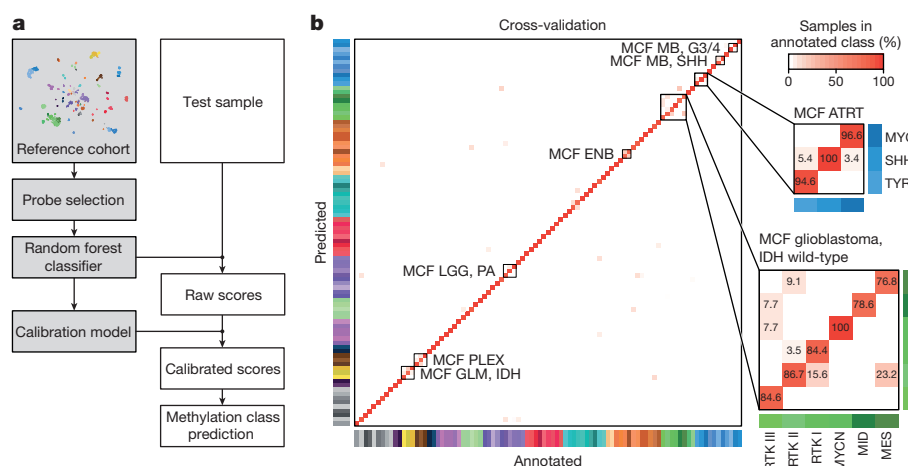


Figure 2 | Development and cross-validation of the DNA methylation-based CNS tumour classifier. **a**, Schematic of principal classifier components (grey) and processing steps for individual test samples (white). The most informative probes are selected for training of the random forest classifier. The classifier produces raw scores that represent the number of decision trees that assign a test sample to a specific methylation class. To enable inter-class comparability, a calibration model is used, which transforms raw scores into calibrated scores. Calibrated

devoted an optimal 'common' calibrated score threshold of ≥ 0.9 (Extended Data Fig. 5d, e). For subclasses within methylation class families, we defined a threshold value of ≥ 0.5 as sufficient for a valid prediction, as long as all family member scores add up to a total score of ≥ 0.9 . Single class specificity and sensitivity for the ≥ 0.9 threshold are provided in Supplementary Table 3.

Clinical implementation

For evaluation of clinical utility, we prospectively analysed a series of 1,155 diagnostic CNS tumours in parallel with standard histopathological analyses (Fig. 3a, b). For 51 cases (4%) the material was not suitable for methylation profiling, mostly because the tumour cell content was too low or because of limited total material. Methylation profiling was performed for the remaining 1,104 samples and the cases were assigned as either 'matching to a defined DNA methylation class' (calibrated score ≥ 0.9) or as 'no match' cases (highest score < 0.9) (for a case-by-case list, see Supplementary Table 4). The investigated cases comprised 64 different histopathological entities from both adult (71%) and paediatric patients (29%). The spectrum of entities was enriched for rare and difficult to diagnose cases received for referral, and therefore did not exactly match the distribution seen in daily routine diagnostic practice. Histopathological evaluation was

performed blinded to DNA methylation profiling results and included standard molecular testing.

In total, 88% of profiled samples ($n = 977$ out of 1,104) matched to an established DNA methylation class with a calibrated classifier score ≥ 0.9 (Fig. 3b). For 838 of these (838 out of 1,104; 76%), results obtained by pathology and DNA methylation profiling were concordant. In 171 of the cases, an unambiguous molecular subgroup could be assigned, which would not have been available based on histopathology evaluation only (for example, molecular subgroups of medulloblastoma and ependymoma, many of which were included in the latest version of the WHO classification of CNS tumours¹).

For the remaining 139 samples with a calibrated classifier score ≥ 0.9 , the DNA methylation class was not in accordance with the pathological diagnosis. These cases were histologically and molecularly re-evaluated, including additional molecular diagnostics (DNA copy-number profiling, targeted gene sequencing, gene panel sequencing²¹ and gene-fusion analysis of a subset of cases, see Supplementary Table 5). This resulted in a revision of the initial histopathological diagnosis in 129 of the 139 cases (12% of all cases; Fig. 4) in favour of the predicted methylation class. In agreement with several recent reports^{16,22,23}, several of these were IDH-wild-type astrocytomas and anaplastic astrocytomas reclassified

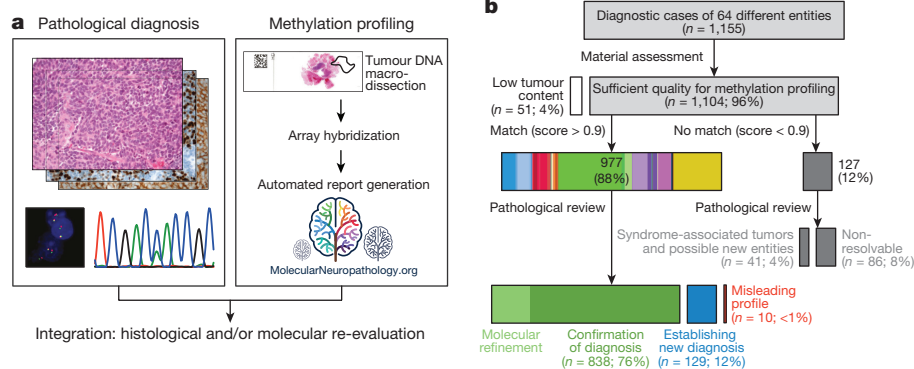


Figure 3 | Implementation of the classifier in diagnostic practice. **a**, Classifier validation by an independent prospective cohort of diagnostic samples. Pathological diagnosis was established by current pathological standard according to the 2016 version of the WHO classification of CNS tumours and compared to classification by methylation profiling.

Cases were categorized as 'confirmation of diagnosis', 'establishing new diagnosis', 'misleading profile' or 'no match to defined class'. **b**, Overview of methylation profiling result from 1,155 diagnostic samples and integration with pathological diagnosis.

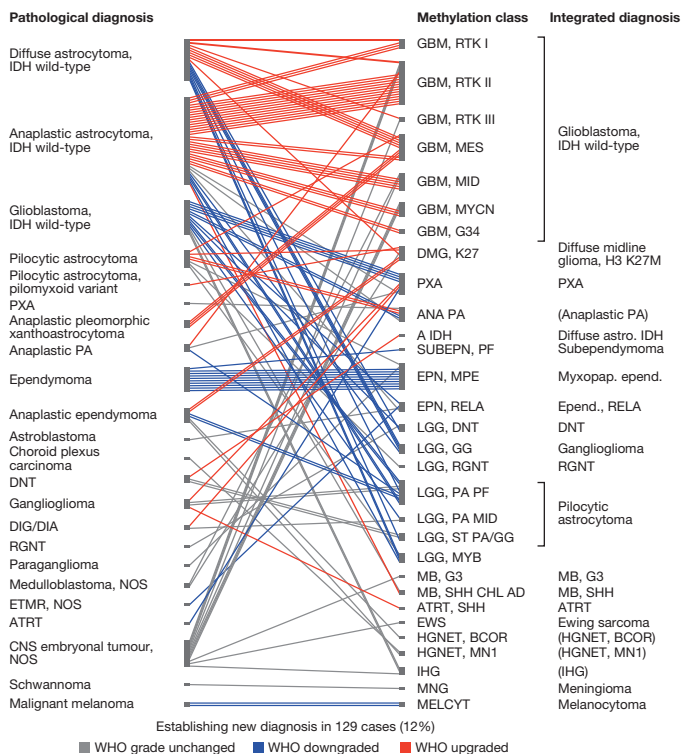


Figure 4 | Reassessment of discrepant cases and establishment of new diagnosis. Discrepancy between pathological diagnosis (left) and methylation profiling (middle) was observed for 139 cases. For 129 cases, histological and molecular reassessment (Supplementary Table 5) resulted in a change in the initial diagnosis with formulation of a new integrated diagnosis (right). For 92 cases, this involved a change in the WHO grading, with both down- (blue) and upgrading (red). Integrated diagnoses in brackets are not recognized as a WHO entity. For methylation class abbreviations see Supplementary Table 1.

as IDH-wild-type glioblastomas. Establishing a new diagnosis had a profound clinical impact: a change in WHO grading was observed in 71% of these cases (92 out of 129), and both upgrading (41%, 53 out of 129) and downgrading (30%, 39 out of 129; Fig. 4). Discrepant results could not be resolved in only 10 cases (<1% of profiled cases), and the histopathological diagnosis was therefore retained.

To substantiate the impact in clinical practice, we contacted five external centres that have started to implement methylation profiling for diagnostic cases using our algorithm. In total, these centres analysed 401 diagnostic cases and in 50 cases (12%) a new diagnosis was established after methylation profiling, very closely recapitulating our rate of reclassification (Extended Data Fig. 6a and Supplementary Table 6). For individual centres, the rate of reclassification varied between 6% and 25%, which was most likely due to differences in the spectrum of investigated cases and more upfront molecular testing by some centres (Extended Data Fig. 6b and Supplementary Table 6).

Twelve per cent of tumours from the prospective cohort (127 out of 1,104) could not be assigned to a DNA methylation class using the rigid calibrated classifier score cutoff of ≥ 0.9 (Fig. 3b). To further clarify the role of these non-classifiable cases, we performed an unsupervised *t*-SNE analysis of the reference cohort together with the diagnostic cohort (Fig. 5a). Using this analysis, we found a high overlap of the classifiable cases with the reference cohort, whereas non-classifiable cases frequently fell in the periphery of the reference classes or were even completely separate from these and frequently grouped with other non-classifiable cases (Fig. 5a). This may indicate that such cases represent rare novel molecular entities that have not been recognized previously. An example for a likely novel CNS tumour entity is shown in Fig. 5b, c.

Technical and inter-laboratory testing

The technical robustness of the random forest classifier was investigated by inter-laboratory comparison. Results of two independent laboratories (starting from DNA extraction) were highly correlated, with only 2 out of 53 samples (4%) showing a classifier score slightly lower than 0.9 in one of the centres, whereas all other cases were classified identically (Extended Data Fig. 7a). Calculation of copy-number profiles was also stable across laboratories (Extended Data Fig. 7b). To ascertain forward compatibility with developing technologies, we further used the random forest classifier to investigate newer Infinium MethylationEPIC BeadChip DNA methylation arrays and high-coverage whole-genome bisulphite sequencing data. For all 16 samples from different CNS tumours that were profiled on both array platforms, raw scores (Extended Data Fig. 7c) and calibrated scores (data not shown) were highly correlated and running them through the classifying algorithm resulted in identical class-assignment for each case. Furthermore, for all 50 high-coverage whole-genome bisulphite sequencing samples (11 different CNS tumour entities), the highest prediction score was for the same class as with the Infinium HumanMethylation450K BeadChip array, suggesting that our approach is applicable to different DNA methylation profiling techniques with only slight adaptations (Extended Data Fig. 7d).

Global dissemination of the platform

To ensure unrestricted community access to our classification system, we created a free web platform for data upload, automatic normalization, random forest classification and PDF report generation (<https://www.molecularneuropathology.org>). DNA copy-number profiles²⁴ and O⁶-methylguanine-DNA-methyltransferase (*MGMT*) promoter methylation status²⁵ are additionally provided, since they can be generated from the same data source—thus having the potential of replacing several time- and cost-intensive single-gene tests. A representative website report is shown in Extended Data Fig. 8. When data are uploaded, the data provider can choose to give consent that the data are allowed to be used for further classifier development. We expect that this web platform can therefore act as a hub for a worldwide cooperative network to continuously identify and track rare tumour classes so that they can eventually be added to the catalogue of known human cancers. In the first year after launch in December 2016, over 4,500 cases have been uploaded from over 15 participating centres. New biological insights are also likely to be gained based on the interrelationships of tumour classes, and by closer examination of how differential DNA methylation affects tumour biology.

Discussion

We here demonstrate that DNA methylation-based CNS tumour classification using a comprehensive machine-learning approach is a valuable asset for clinical decision-making. In particular, the high level of standardization has great promise to reduce the substantial inter-observer variability observed in current CNS tumour diagnostics. Furthermore, in contrast to traditional pathology, where there is a pressure to assign all tumours to a described entity even for atypical or challenging cases, the objective measure that we provide here allows for 'no match' to a defined class. This information can also be of substantial value to highlight that a tumour is not a typical example of a given differential diagnosis, and may rather belong to a rarer, currently undefined class. We defined five categories of methylation classes that have different clinical implications. Category 1 can be directly translated to WHO entities. Category 2 represents subclasses of WHO entities. For all but ependymal tumours, subclassification currently has little clinical consequence and a translation back to the WHO class may be appropriate for clinical purposes. Category 3 reflects the fact that WHO grading cannot be fully recapitulated by methylation profiling for several classes. Further data are required to assess whether the methylation classes of this category may provide a more robust means of prognostication than histology alone, as has been demonstrated for

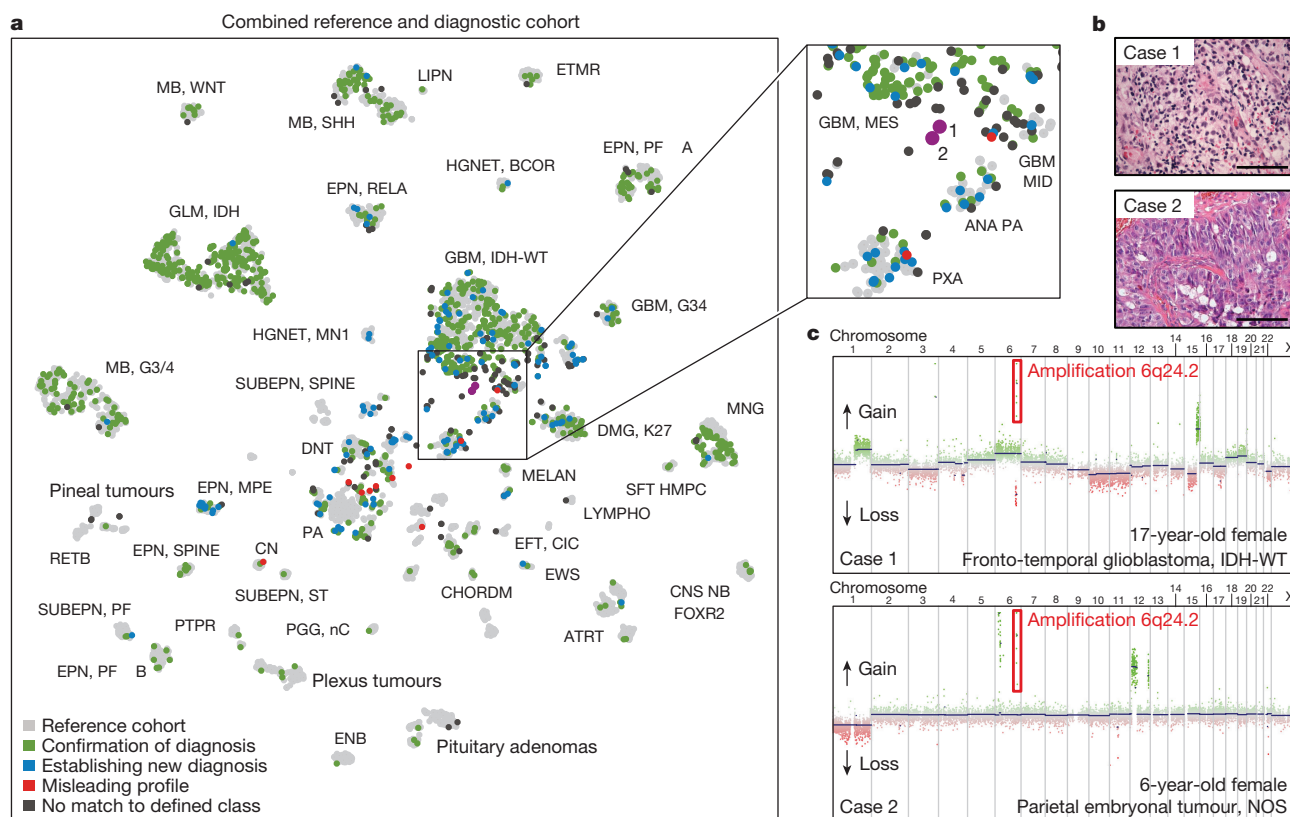


Figure 5 | DNA methylation-based identification of potential new CNS tumour entities. **a**, Unsupervised clustering of the combined reference ($n = 2,801$, grey) and diagnostic cohort ($n = 1,104$, coloured) using t -SNE dimensionality reduction. Abbreviated names indicate the reference cohort classes as in Fig. 1. The diagnostic samples are colour-coded as confirmation of diagnosis ($n = 838$, green), establishing new diagnosis ($n = 129$, blue), misleading profile ($n = 10$, red) and no match to defined class ($n = 127$, dark grey). The matching (green) and reclassified (blue) cases show high overlap with the reference cases. The non-classifiable (black) and the misleading (red) cases frequently fall in the periphery of the reference classes or are completely separate of these. The magnified area (right) highlights two non-classifiable cases (here in magenta for easier identification) that group together in the t -SNE representation.

several other classes^{4,9,11}. In category 4, the WHO entity boundaries are not identical to the boundaries of the methylation classes. Until additional data on the exact boundaries become available, this category should be critically discussed in the clinical context and orthogonal testing should be undertaken whenever possible. Category 5 represents putative new entities that are currently not recognized by the WHO, and although limited data on these cases are currently available, the biological rationale for a novel class was considered to be strong.

A study in which reference pathology and molecular diagnostics, including DNA methylation profiling, are blinded for each other's results is currently ongoing for all childhood brain tumours diagnosed in Germany to objectivize the potential effect of re-classification on patient outcome (Molecular Neuropathology 2.0; <http://www.kitz-heidelberg.de/molecular-diagnostics>), with results due over the next few years.

A uniform implementation of the classification algorithm holds great promise for standardization of tumour diagnostics across centres and across clinical trials. Furthermore, the digital nature of methylation data facilitates easy exchange and will allow aggregation of extensive tumour libraries. This will probably result in the detection of exceptionally rare tumour classes and a continued refinement of classifiers. Inclusion of new classes will allow a prompt translation into diagnostic practice, almost certainly resulting in a more dynamic tumour classification. In our experience, adaptation of this technique in diagnostic laboratories

is relatively straightforward. Extended Data Fig. 9 summarizes a sample workflow for diagnostic implementation. We expect that the principle of using DNA methylation signatures as part of a combined histology and molecular tumour classification will improve diagnostic accuracy not only in neuropathology, but will also serve as a blueprint in other fields of tumour pathology.

Online Content Methods, along with any additional Extended Data display items and Source Data, are available in the online version of the paper; references unique to these sections appear only in the online paper.

Received 5 May 2017; accepted 13 February 2018.

Published online 14 March 2018.

- Louis, D. N., Ohgaki, H., Wiestler, O. D. & Cavenee, W. K. *WHO Classification of Tumours of the Central Nervous System* revised 4th edn (IARC, 2016).
- van den Bent, M. J. Interobserver variation of the histopathological diagnosis in clinical trials on glioma: a clinician's perspective. *Acta Neuropathol.* **120**, 297–304 (2010).
- Ellison, D. W. *et al.* Histopathological grading of pediatric ependymoma: reproducibility and clinical relevance in European trial cohorts. *J. Negat. Results Biomed.* **10**, 7 (2011).
- Sturm, D. *et al.* New brain tumor entities emerge from molecular classification of CNS-PNETs. *Cell* **164**, 1060–1072 (2016).
- Fernandez, A. F. *et al.* A DNA methylation fingerprint of 1628 human samples. *Genome Res.* **22**, 407–419 (2012).
- Hovestadt, V. *et al.* Decoding the regulatory landscape of medulloblastoma using DNA methylation sequencing. *Nature* **510**, 537–541 (2014).

7. Moran, S. *et al.* Epigenetic profiling to classify cancer of unknown primary: a multicentre, retrospective analysis. *Lancet Oncol.* **17**, 1386–1395 (2016).
8. Hovestadt, V. *et al.* Robust molecular subgrouping and copy-number profiling of medulloblastoma from small amounts of archival tumour material using high-density DNA methylation arrays. *Acta Neuropathol.* **125**, 913–916 (2013).
9. Sturm, D. *et al.* Hotspot mutations in *H3F3A* and *IDH1* define distinct epigenetic and biological subgroups of glioblastoma. *Cancer Cell* **22**, 425–437 (2012).
10. Reuss, D. E. *et al.* Adult IDH wild type astrocytomas biologically and clinically resolve into other tumor entities. *Acta Neuropathol.* **130**, 407–417 (2015).
11. Pajtler, K. W. *et al.* Molecular classification of ependymal tumors across all CNS compartments, histopathological grades, and age groups. *Cancer Cell* **27**, 728–743 (2015).
12. Lambert, S. R. *et al.* Differential expression and methylation of brain developmental genes define location-specific subsets of pilocytic astrocytoma. *Acta Neuropathol.* **126**, 291–301 (2013).
13. Thomas, C. *et al.* Methylation profiling of choroid plexus tumors reveals 3 clinically distinct subgroups. *Neuro-oncol.* **18**, 790–796 (2016).
14. Mack, S. C. *et al.* Epigenomic alterations define lethal CIMP-positive ependymomas of infancy. *Nature* **506**, 445–450 (2014).
15. Johann, P. D. *et al.* Atypical teratoid/rhabdoid tumors are comprised of three epigenetic subgroups with distinct enhancer landscapes. *Cancer Cell* **29**, 379–393 (2016).
16. Wiestler, B. *et al.* Integrated DNA methylation and copy-number profiling identify three clinically and biologically relevant groups of anaplastic glioma. *Acta Neuropathol.* **128**, 561–571 (2014).
17. Van Der Maaten, L. & Hinton, G. H. Visualizing data using *t*-SNE. *J. Mach. Learn. Res.* **9**, 2579–2605 (2008).
18. Ceccarelli, M. *et al.* Molecular profiling reveals biologically discrete subsets and pathways of progression in diffuse glioma. *Cell* **164**, 550–563 (2016).
19. Breiman, L. Random forests. *Mach. Learn.* **45**, 5–32 (2001).
20. Sokolova, M. & Lapalme, G. A systematic analysis of performance measures for classification tasks. *Inf. Process. Manage.* **45**, 427–437 (2009).
21. Sahm, F. *et al.* Next-generation sequencing in routine brain tumor diagnostics enables an integrated diagnosis and identifies actionable targets. *Acta Neuropathol.* **131**, 903–910 (2016).
22. Weller, M. *et al.* Molecular classification of diffuse cerebral WHO grade II/III gliomas using genome- and transcriptome-wide profiling improves stratification of prognostically distinct patient groups. *Acta Neuropathol.* **129**, 679–693 (2015).
23. The Cancer Genome Atlas Research Network. Comprehensive, integrative genomic analysis of diffuse lower-grade gliomas. *N. Engl. J. Med.* **372**, 2481–2498 (2015).
24. Hovestadt, V. & Zapatka, M. conumee: enhanced copy-number variation analysis using Illumina methylation arrays. v.1.4.2 R package v.0.99.4 <http://www.bioconductor.org/packages/release/bioc/html/conumee.html> (2015).
25. Bady, P., Delorenzi, M. & Hegi, M. E. Sensitivity analysis of the MGMT-STP27 model and impact of genetic and epigenetic context to predict the MGMT methylation status in gliomas and other tumors. *J. Mol. Diagn.* **18**, 350–361 (2016).

Supplementary Information is available in the online version of the paper.

Acknowledgements We thank U. Lass, A. Habel, I. Oezen for technical and administrative support, the Microarray unit of the Genomics and Proteomics Core Facility (DKFZ) for methylation services, the German Glioma Network and the Neuroonkologische Arbeitsgemeinschaft for sharing their data. This research was supported by the DKFZ-Heidelberg Center for Personalized Oncology (DKFZ-HIPO_036), the German Childhood Cancer Foundation (DKS 2015.01), an Illumina Medical Research Grant, the DKTK joint funding project 'Next Generation Molecular Diagnostics of Malignant Gliomas', the A Kids' Brain Tumour Cure (PLGA) Foundation, the Brain Tumour Charity (UK) for the Everest Centre for Paediatric Low-Grade Brain Tumour Research, the Friedberg Charitable Foundation and the Sohn Conference Foundation (to M. Snuderl and M. Karajannis), the RKA-Förderpool (Project 37) and Stichting Kinderen Kankervrij and Stichting AMC Foundation (to E. Aronica), NIH/NCI 5T32CA163185 (to A.O.), NIH/NCI Cancer Center Support Grant P30 CA008748 to MSKCC, the Luxembourg National Research Fond (FNR PEARL P16/BM/11192868 to M.M.) and the National Institute of Health Research (NIHR) UCLH/UCL Biomedical Research Centre (S.Bra.).

Author Contributions D.C. and D.T.W.J. composed the reference cohort and defined methylation classes; M.Si. and V.Ho. developed and technically validated the classification algorithm; D.Sc. developed the classification website; D.C., D.T.W.J., M.Si., A.Ben., V.Ho., D.Sc., D.Sti., M.Z., A.v.D. and S.M.P. developed additional methodology and software; D.C., D.T.W.J., M.Si., D.Stu., C.Ko., F.Sa., L.C., D.E.R., A.Kr., A.K.W., K.H., L.S., P.N.H., K.H.P., J.Schi., G.R., M.Pri., W.B., F.Se., H.W., T.M., O.W., S.Bre., M.S.-R., D.H., A.Ku., C.M.K., H.L.M., S.Ru., K.v.H., M.C.F., A.Gn., G.F., S.T., G.C., C.-M.M., M.G., T.P., M.B., J.D., M.Pl., A.U., W.W., M.M., C.Har., C.H.-M., M.H., A.Kor., A.v.D. and S.M.P. performed the prospective cohort analysis; P.N.H., K.H.P., H.D., B.K.G., J.H., S.F., P.W., Z.J., T.A., S.Bra. generated and collected the external centre data; K.W.P., A.O., N.W.E., A.K.B., R.C., A.Hö., E.H., R.Be., J.Schi., O.S., K.W., P.V., M.Pa., P.T., D.L., E.A., F.G., E.R., W.S., C.G., F.J.R., A.Bec., M.Pre., C.Hab., R.Bj., J.C., M.F., M.D., S.H., V.Ha., S.Ro., J.R.H., P.K., B.W.K., M.L., B.L., C.M., R.K., Z.K., F.H., A.Koc., A.J., C.Ke., H.M., W.M., U.P., M.Pri., N.G.G., P.H.D., A.P., C.J., T.S.J., B.R., T.P., J.Schr., G.S., M.Wes., G.R., P.W., M.Wel., V.P.C., I.B., A.Hu., N.J., P.A.N., W.P., A.Ga., G.W.R., M.D.T., M.R., M.A.K., M.M., C.Har., K.A., U.S., R.Bu., P.L., M.K., C.H.-M., D.W.E., M.H., S.Bra., A.Kor., A.v.D. and S.M.P. provided reference cohort material and data; K.L., M.B.-H., M.Sc. and R.F. performed methylation profiling; J.Se., K.K., A.T., M.K. and M.Sn. performed technical validation experiments; A.v.D. and S.M.P. supervised the project. The manuscript underwent an internal collaboration-wide review process. All authors approved the final version of the manuscript.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare competing financial interests: details are available in the online version of the paper. Readers are welcome to comment on the online version of the paper. Publisher's note: Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations. Correspondence and requests for materials should be addressed to S.M.P. (s.pfister@dkfz.de) or A.v.D. (Andreas.vonDeimling@med.uni-heidelberg.de).

Reviewer Information Nature thanks S. Pomeroy, M. L. Suva, R. Verhaak and S. Yip for their contribution to the peer review of this work.

David Capper^{1,2,3,4*}, David T. W. Jones^{5,6*}, Martin Sill^{5,6,7*}, Volker Hove stad^{8,9*}, Daniel Schrimpf^{1,2}, Dominik Sturm^{5,6,9}, Christian Koelsche^{1,2}, Felix Sahm^{1,2}, Lukas Chavez^{5,6}, David E. Reuss^{1,2}, Annekatrin Kratz^{1,2}, Annika K. Wefers^{1,2}, Kristian Huang^{1,2}, Kristian W. Pajtl^{5,6,9}, Leonille Schweizer^{1,3}, Damian Stichel^{1,2}, Adriana Olar^{10,11,12}, Nils W. Engel^{13,14}, Kerstin Lindenberg², Patrick N. Harter^{15,16}, Anne K. Braczynski^{15,16}, Karl H. Plate^{15,16}, Hildegard Dohmen¹⁷, Boyan K. Garvalov¹⁷, Roland Coras¹⁸, Annett Hölsken¹⁸, Ekkehard Hewer¹⁹, Melanie Bewerunge-Hudler²⁰, Matthias Schick²⁰, Roger Fischer²⁰, Rudi Beschoner²¹, Jens Schittenhelm²¹, Ori Staszewski²², Khalida Wani²³, Pascale Varlet²⁴, Melanie Pages²⁴, Petra Temming²⁵, Dietmar Lohmann²⁶, Florian Selt^{5,9,27}, Hendrik Witt^{5,6,9}, Till Milde^{5,9,27}, Olaf Witt^{5,9,27}, Eleonora Aronica^{28,29,30}, Felice Giangaspero^{31,32}, Elisabeth Rushing³³, Wolfram Scheurlen³⁴, Christoph Geisenberger^{35,36}, Fausto J. Rodriguez³⁷, Albert Becker³⁸, Matthias Preusser³⁹, Christine Haberler⁴⁰, Rolf Bjerkvig^{41,42}, Jane Cryan⁴³, Michael Farrell⁴³, Martina Deckert⁴⁴, Jürgen Hench⁴⁵, Stephan Frank⁴⁵, Jonathan Serrano⁴⁶, Kasthuri Kannan⁴⁶, Aristotelis Tsigros⁴⁶, Wolfgang Brück⁴⁷, Silvia Hofer⁴⁸, Stefanie Brehmer⁴⁹, Marcel Seiz-Rosenhagen⁴⁹, Daniel Hänggi⁴⁹, Volkmar Hans^{50,51}, Stephanie Rozsnoki⁵², Jordan R. Hansford^{53,54,55}, Patricia Kohlhor⁵⁶, Bjarne W. Kristensen⁵⁷, Matt Lechner⁵⁸, Beatriz Lopes⁵⁹, Christian Mawrin⁶⁰, Ralf Ketter⁶¹, Andreas Kulozik^{5,9}, Ziad Khatib⁶², Frank Heppner^{3,63,64}, Arend Koch³, Anne Jouvét⁶⁵, Catherine Keohane⁶⁶, Helmut Mühleisen⁶⁷, Wolf Mueller⁶⁸, Ute Pohl⁶⁹, Marco Prinz^{22,70}, Axel Benner⁷, Marc Zapatka⁸, Nicholas G. Gottardo^{71,72,73}, Pablo Hernáiz Driever⁷⁴, Christof M. Kramm⁷⁵, Hermann L. Müller⁷⁶, Stefan Rutkowski⁷⁷, Katja von Hoff^{74,77}, Michael C. Frühwald⁷⁸, Astrid Gnekow⁷⁸, Gudrun Fleischhack²⁵, Stephan Tippelt²⁵, Gabriele Calaminus⁷⁹, Camelia-Maria Monoranu⁸⁰, Arie Perry⁸¹, Chris Jones⁸², Thomas S. Jacques⁸³, Bernhard Radlwimmer⁸, Marco Gessi³⁸, Torsten Pietsch³⁸, Johannes Schramm⁸⁴, Gabriele Schackert⁸⁵, Manfred Westphal⁸⁶, Guido Reifenberger^{87,88}, Pieter Wesseling^{89,90}, Michael Weller⁹¹, Vincent Peter Collins⁹², Ingmar Blümcke¹⁸, Martin Bendszus⁹³, Jürgen Debus⁹⁴, Annie Huang⁹⁵, Nada Jabado⁹⁶, Paul A. Northcott⁹⁷, Werner Paulus⁵², Amar Gajjar⁹⁸, Giles W. Robinson⁹⁸, Michael D. Taylor⁹⁹, Zane Jaunmuktane^{100,101,102}, Marina Ryzhova¹⁰³, Michael Platten¹⁰⁴, Andreas Unterberg³⁵, Wolfgang Wick¹⁰⁵, Matthias A. Karajannis¹⁰⁶, Michel Mittelbronn^{15,16,107,108,109,110}, Till Acker¹⁷, Christian Hartmann¹¹¹, Kenneth Aldape¹¹², Ulrich Schüller^{14,113,114,115}, Rolf Buslei^{18,116}, Peter Lichter⁸, Marcel Kool^{5,6}, Christel Herold-Mende³⁵, David W. Ellison¹¹⁷, Martin Hasselblatt⁵², Matija Snuderl¹¹⁸, Sebastian Brandner^{100,102}, Andrey Korshunov^{1,2}, Andreas von Deimling^{1,2} & Stefan M. Pfister^{5,6,9}§

¹Department of Neuropathology, University Hospital Heidelberg, Heidelberg, Germany. ²Clinical Cooperation Unit Neuropathology, German Cancer Consortium (DKTK), German Cancer Research Center (DKFZ), Heidelberg, Germany. ³Charité — Universitätsmedizin Berlin, corporate member of Freie Universität Berlin, Humboldt-Universität zu Berlin, and Berlin Institute of Health, Department of Neuropathology, Berlin, Germany. ⁴German Cancer Consortium (DKTK), Partner Site Berlin, German Cancer Research Center (DKFZ), Heidelberg, Germany. ⁵Hopp Children's Cancer Center at the NCT Heidelberg (KITZ), Heidelberg, Germany. ⁶Division of Pediatric Neurooncology, German Cancer Consortium (DKTK), German Cancer Research Center (DKFZ), Heidelberg, Germany. ⁷Division of Biostatistics, German Cancer Research Center (DKFZ), Heidelberg, Germany. ⁸Division of Molecular Genetics, German Cancer Research Center (DKFZ), Heidelberg, Germany. ⁹Department of Pediatric Oncology, Hematology and Immunology, University Hospital Heidelberg, Heidelberg, Germany. ¹⁰Department of Pathology and Laboratory Medicine, Medical University of South Carolina, Charleston, South Carolina 29425, USA. ¹¹Department of Neurosurgery, Medical University of South Carolina, Charleston, South Carolina 29425, USA. ¹²Hollings Cancer Center, Charleston, South Carolina 29425, USA. ¹³Department of Oncology and Hematology with Sections Bone Marrow Transplant and Pneumology, Hubertus Wald Tumorzentrum/University Cancer Center Hamburg, University Medical Center Hamburg, Hamburg, Germany. ¹⁴Center for Neuropathology and Prion Research, Ludwig-Maximilians-University, Munich, Germany. ¹⁵Institute of Neurology (Edinger Institute), Goethe-University Frankfurt am Main, Frankfurt am Main, Germany. ¹⁶German Cancer Consortium (DKTK), Partner Site Frankfurt/Mainz, Frankfurt am Main, German Cancer Research Center (DKFZ) Heidelberg, Germany. ¹⁷Institute of Neuropathology, University of Giessen, Giessen, Germany. ¹⁸Neuropathological Institute, University Hospital Erlangen, Friedrich Alexander University Erlangen-Nuremberg, Erlangen, Germany. ¹⁹Institute of Pathology, University of Bern, Bern, Switzerland. ²⁰Genomics and Proteomics Core Facility, German Cancer Research Center (DKFZ), Heidelberg, Germany. ²¹Institute of Pathology and Neuropathology, Department of Neuropathology, University Hospital Tübingen, Tübingen, Germany. ²²Institute of Neuropathology, Medical Center—University of Freiburg, Faculty of Medicine, University of Freiburg, Germany. ²³Department of Translational Molecular Pathology, University of Texas MD Anderson Cancer Center, Houston, Texas 77030, USA. ²⁴Department of Neuropathology, Centre Hospitalier Sainte Anne, Paris, France. ²⁵Pediatrics III, Pediatric Oncology and Hematology, University Hospital Essen, Essen, Germany. ²⁶Eye Cancer Research Group, Faculty of Medicine, University of Duisburg-Essen, Essen, Germany. ²⁷CCU Pediatric Oncology (G340), German Cancer Research Center (DKFZ) and German Cancer Consortium (DKTK), Heidelberg, Germany. ²⁸Department of (Neuro) Pathology, Academic Medisch Centrum (AMC), University of Amsterdam, Amsterdam, The Netherlands. ²⁹Swammerdam Institute for Life Sciences, Center for Neuroscience, University of Amsterdam, Amsterdam, The Netherlands. ³⁰Stichting Epilepsie Instellingen Nederland (SEIN), Amsterdam, The Netherlands. ³¹Department of Radiological, Oncological and Anatomopathological Sciences, Sapienza University, Rome, Italy. ³²IRCCS Neuromed, Pozzilli, Italy. ³³Department of Neuropathology, University Hospital Zurich, Zurich, Switzerland. ³⁴Cnopf'sche

Kinderklinik Nuremberg, Nuremberg, Germany. ³⁵Department of Neurosurgery, Heidelberg University Hospital, Heidelberg, Germany. ³⁶Hubrecht Institute-KNAW (Royal Netherlands Academy of Arts and Sciences), Uppsalalaan 8, 3584 CT Utrecht, The Netherlands. ³⁷Division of Neuropathology of the Johns Hopkins University School of Medicine, Baltimore, Maryland, USA. ³⁸Department of Neuropathology, University of Bonn, Bonn, Germany. ³⁹Department of Medicine I, Comprehensive Cancer Center Vienna, CNS Unit (CCC-CNS), Medical University of Vienna, Vienna, Austria. ⁴⁰Institute of Neurology, Medical University of Vienna, Vienna, Austria. ⁴¹Department of Biomedicine, University of Bergen, Bergen, Norway. ⁴²NORLUX Neuro-Oncology Laboratory, Department of Oncology, Luxembourg Institute of Health, Luxembourg, Luxembourg. ⁴³Department of Neuropathology, Beaumont Hospital, Dublin, Ireland. ⁴⁴Department of Neuropathology, University Hospital of Cologne, Cologne, Germany. ⁴⁵Department of Neuropathology, Institute of Pathology, Basel University Hospital, Basel, Switzerland. ⁴⁶NYU Langone Medical Center, New York, New York, USA. ⁴⁷Institute of Neuropathology, University Medical Center Göttingen, Göttingen, Germany. ⁴⁸Division of Oncology, Luzerner Kantonsspital, Luzern, Switzerland. ⁴⁹Department of Neurosurgery, University Medical Center Mannheim, University of Heidelberg, Mannheim, Germany. ⁵⁰Institut für Neuropathologie, Evangelisches Krankenhaus Bielefeld gGmbH, Bielefeld, Germany. ⁵¹Institut für Neuropathologie, Universitätsklinikum Essen, Essen, Germany. ⁵²Institute of Neuropathology, University Hospital Münster, Münster, Germany. ⁵³Children's Cancer Centre, Royal Children's Hospital, University of Melbourne, Melbourne, Victoria, Australia. ⁵⁴Murdoch Children's Research Institute, University of Melbourne, Melbourne, Victoria, Australia. ⁵⁵Department of Pediatrics, University of Melbourne, Melbourne, Victoria, Australia. ⁵⁶Institute for Pathology, Katharinenhospital Stuttgart, Stuttgart, Germany. ⁵⁷Department of Pathology, Odense University Hospital, Department of Clinical Research, University of Southern Denmark, Odense, Denmark. ⁵⁸University College London Cancer Institute and University College London Hospitals, London, UK. ⁵⁹Department of Pathology, University of Virginia, Charlottesville, Virginia, USA. ⁶⁰Institute of Neuropathology, Otto-von-Guericke-University, Magdeburg, Germany. ⁶¹Department of Neurosurgery, University Hospital Saarland, Homburg, Saar, Germany. ⁶²Nicklaus Children's Hospital Brain Institute, Miami, Florida 33155, USA. ⁶³Cluster of Excellence, NeuroCure, Berlin, Germany. ⁶⁴Berlin Institute of Health (BIH), Berlin, Germany. ⁶⁵Département de Pathologie et Neuropathologie, Hôpital Neurologique, Hospices Civils de Lyon, Lyon, France. ⁶⁶Department of Neuropathology, Cork University Hospital, Cork, Ireland. ⁶⁷Department of Pathology, Ludwigsburg Hospital, Ludwigsburg, Germany. ⁶⁸Department of Neuropathology, Leipzig University, Leipzig, Germany. ⁶⁹Department of Cellular Pathology, Queen's Hospital, Romford, UK. ⁷⁰BIOSS Centre for Biological Signalling Studies, University of Freiburg, Freiburg, Germany. ⁷¹Department of Pediatric Oncology and Haematology, Princess Margaret Hospital for Children, GPO Box D184, Perth, Western Australia 6840, Australia. ⁷²Telethon Kids Institute, University of Western Australia, PO Box 855, Perth, Western Australia 6872, Australia. ⁷³School of Paediatrics and Child Health, University of Western Australia, GPO Box D184, Perth, Western Australia 6840, Australia. ⁷⁴Department of Pediatric Oncology/Hematology, Charité-Universitätsmedizin Berlin, Berlin, Germany. ⁷⁵Division of Pediatric Hematology and Oncology, University Medical Center Göttingen, Göttingen, Germany. ⁷⁶Department of Pediatrics and Pediatric Hematology/Oncology, Klinikum Oldenburg AöR, Medical Campus University Oldenburg, 26133 Oldenburg, Germany. ⁷⁷Department for Pediatric Hematology and Oncology, University Hospital Hamburg-Eppendorf, Hamburg, Germany. ⁷⁸Children's Hospital Augsburg, Swabian Children's Cancer Centre, Augsburg, Germany. ⁷⁹Department of Pediatric Hematology/Oncology, University of Bonn Medical Center, Bonn, Germany. ⁸⁰Department of Neuropathology, Institute of Pathology, Comprehensive Cancer Center (CCC) Mainfranken, University of Würzburg, Würzburg, Germany. ⁸¹Department of Pathology, University of California San Francisco, San Francisco, California, USA. ⁸²Division of Molecular Pathology, Institute of Cancer Research, London, UK. ⁸³Developmental Biology and Cancer Programme, UCL Great Ormond Street Institute of Child Health and Histopathology Department, Great Ormond Street Hospital for Children NHS Foundation Trust, London, UK. ⁸⁴Medical Faculty, University of Bonn Medical School, Bonn, Germany. ⁸⁵Department of Neurosurgery, University Hospital Carl Gustav Carus, Technische Universität Dresden, Dresden, Germany. ⁸⁶Department of Neurosurgery, University Hospital Hamburg-Eppendorf, Hamburg, Germany. ⁸⁷Department of Neuropathology, Heinrich Heine University Düsseldorf, Düsseldorf, Germany. ⁸⁸German Cancer Consortium (DKTK), partner site Essen/Düsseldorf, German Cancer Research Center (DKFZ), Heidelberg, Germany. ⁸⁹Department of Pathology, Princess Máxima Center for Pediatric Oncology and University Medical Center Utrecht, Utrecht, The Netherlands. ⁹⁰Department of Pathology, VU University Medical Center, Amsterdam, The Netherlands. ⁹¹Department of Neurology, University Hospital and University of Zurich, Zurich, Switzerland. ⁹²Department of Pathology, Division of Molecular Histopathology, University of Cambridge, Cambridge, UK. ⁹³Department of Neuroradiology, Heidelberg University Hospital, Heidelberg, Germany. ⁹⁴Department of Radiation Oncology, Heidelberg University Hospital, Heidelberg, Germany. ⁹⁵Department of Pediatrics, Laboratory Medicine and Pathobiology, University of Toronto, Toronto, Ontario, Canada. ⁹⁶Division of Hematology/Oncology, McGill University, Montreal, Quebec, Canada. ⁹⁷Department of Developmental Neurobiology, St Jude Children's Research Hospital, Memphis, Tennessee, USA. ⁹⁸Department of Oncology, St Jude Children's Research Hospital, Memphis, Tennessee, USA. ⁹⁹Division of Neurosurgery, Arthur and Sonia Labatt Brain Tumor Research Centre, Hospital for Sick Children, University of Toronto, Toronto, Ontario, Canada. ¹⁰⁰Division of Neuropathology, UCL Hospitals, Institute of Neurology, University College London, Queen Square, WC1N 3BG London, UK. ¹⁰¹Department of Molecular Neuroscience, Institute of Neurology, University College London, Queen Square, WC1N 3BG London, UK. ¹⁰²Department of Neurodegeneration, Institute of Neurology, University College London, Queen Square, WC1N 3BG London, UK. ¹⁰³NN Burdenko Neurosurgical Institute, Moscow, Russia. ¹⁰⁴Department of Neurology, Universitätsmedizin Mannheim, Medical Faculty Mannheim, Heidelberg University, Mannheim, Germany. ¹⁰⁵Department of Neurology, Heidelberg University Hospital, Heidelberg, Germany. ¹⁰⁶Department of Pediatrics, Memorial Sloan Kettering Cancer Center, New York, New York, USA. ¹⁰⁷NORLUX Neuro-Oncology Laboratory, Luxembourg Institute of Health (LIH), Luxembourg, Luxembourg. ¹⁰⁸Luxembourg Centre for Systems Biomedicine (LCSB), University

of Luxembourg, Luxembourg, Luxembourg. ¹⁰⁹Laboratoire national de santé (LNS), Dudelange, Luxembourg. ¹¹⁰Luxembourg Centre of Neuropathology (LCNP), Luxembourg, Luxembourg. ¹¹¹Department of Neuropathology, Hannover Medical School (MHH), Hannover, Germany. ¹¹²Department of Pathology, University of Toronto, Toronto, Ontario, Canada. ¹¹³Institute of Neuropathology, University Medical Center, Hamburg-Eppendorf, Hamburg, Germany. ¹¹⁴Research Institute Children's Cancer Center, Hamburg, Germany. ¹¹⁵Department of Pediatric Hematology and Oncology, University Medical Center, Hamburg-Eppendorf, Hamburg, Germany. ¹¹⁶Section Neuropathology, Institute of Pathology, Sozialstiftung Bamberg, Klinikum

am Bruderwald, Bamberg, Germany. ¹¹⁷Department of Pathology, St Jude Children's Research Hospital, Memphis, Tennessee, USA. ¹¹⁸Division of Neuropathology, Department of Pathology, NYU Langone Medical Center, New York, New York, USA. †Present addresses: Department of Pathology and Center for Cancer Research, Massachusetts General Hospital and Harvard Medical School, Boston, Massachusetts, USA (V.Ho.); Broad Institute of MIT and Harvard, Cambridge, Massachusetts, USA (V.Ho.). *These authors contributed equally to this work. §These authors jointly supervised this work.

METHODS

Data reporting. No statistical methods were used to predetermine sample size. The experiments were not randomized and the investigators were not blinded to allocation during experiments and outcome assessment.

Patient material. Patient material and clinical data of the retrospective reference cohort (total, $n = 2,801$) were obtained from the National Center for Tumour Diseases (NCT) in Heidelberg and supplemented with samples from additional centres (Supplementary Table 2) according to protocols approved by the Institutional Review Boards of the appropriate centres with written consent obtained from each patient. Tumours were histopathologically re-assessed according to the current WHO classification¹. Areas with highest tumour cell content ($\geq 70\%$) were selected for DNA extraction. Subsets of the reference cohort have been previously published^{4,6,9–16,26–33}. Additional patient characteristics can be found in Supplementary Table 2. The prospectively assessed clinical cohort was analysed as part of the National Center for Tumour Diseases Precision Oncology Program according to procedures approved by the Institutional Review Board at the Medical Faculty Heidelberg. All patients gave written consent for diagnostic procedures, comprising molecular testing, including methylation profiling. For all the above human research participants, all relevant ethical regulations were followed. Additional patient characteristics can be found in Supplementary Table 4. Details of the online-analysed cohort of the five additional centres can be found in Supplementary Table 6. Usage of the data was according to protocols approved by the Institutional Review Boards of the University of Basel, Frankfurt am Main University Hospital, University Medical Center Utrecht and Princess Máxima Center for Paediatric Oncology Utrecht, Giessen University Hospital and University College London Hospitals.

Data generation, processing and random forest classifier generation. Samples were analysed using Illumina Infinium HumanMethylation450 BeadChip (450k) arrays according to the manufacturer's instructions. To investigate stability across platforms a selection of samples were additionally assessed using the successor Methylation BeadChip (EPIC) array or whole-genome bisulphite sequencing (WGBS, generated and analysed as described⁶). Array data analysis was performed using R v.3.2.0³⁴, using a number of packages from Bioconductor³⁵ and other repositories. A random forest¹⁹ classifier compatible with both 450k and EPIC platforms was trained, and a calibration model that calculates class probabilities from random forest scores was devised.

Methylation array processing. The 450k array was used to obtain genome-wide DNA methylation profiles for tumour samples and normal control tissues, according to the manufacturer's instructions (Illumina). DNA methylation data were generated at the Genomics and Proteomics Core Facility of the DKFZ (Heidelberg, Germany) and the NYU Langone Medical Center (New York, USA). Data were generated from both freshly frozen and formalin-fixed paraffin-embedded (FFPE) tissue samples. For most freshly frozen samples, >500 ng of DNA was used as input material, whereas 250 ng of DNA was used for most FFPE tissues. On-chip quality metrics of all samples were carefully controlled. Copy-number variation analysis from 450k methylation array data was performed using the *conumee*²⁴ Bioconductor package v.1.3.0. Two sets of 50 control samples displaying a balanced copy-number profile from both male and female donors were used for normalization.

Raw signal intensities were obtained from IDAT files using the *minfi* Bioconductor package v.1.14.0³⁶. Each sample was individually normalized by performing a background correction (shifting of the 5% percentile of negative control probe intensities to 0) and a dye-bias correction (scaling of the mean of normalization control probe intensities to 10,000) for both colour channels. Subsequently, a correction for the type of material tissue (FFPE or frozen) was performed by fitting univariate, linear models to the log₂-transformed intensity values (*removeBatchEffect* function, *limma* package v.3.24.15). The methylated and unmethylated signals were corrected individually. Estimated batch effects were also used to adjust diagnostic samples or test samples within the cross-validation. Beta values were calculated from the retransformed intensities using an offset of 100 (as recommended by Illumina). To test for possible confounding batch effects within our pre-processed reference cohort dataset (after adjusting for FFPE versus frozen material) we applied the *sva* algorithm^{37,38}. We found no significant surrogate variable (data not shown).

The following filtering criteria were applied: removal of probes targeting the X and Y chromosomes ($n = 11,551$), removal of probes containing a single-nucleotide polymorphism (dbSNP132 Common) within five base pairs of and including the targeted CpG site ($n = 7,998$), probes not mapping uniquely to the human reference genome (hg19) allowing for one mismatch ($n = 3,965$), and probes not included on the Illumina EPIC array ($n = 32,260$). In total, 428,799 probes targeting CpG sites were kept for further analysis.

Unsupervised analysis. Pairwise Pearson correlation was calculated for all 2,801 reference samples by selecting the 32,000 most variably methylated probes

(s.d. > 0.228 ; Extended Data Fig. 1a). The same probes were used for principal component analysis (PCA). For PCA, pairwise probe covariances of centred beta values were calculated. Eigenvalue decomposition was performed using the *eigs* function of the *RSpectra* package version 0.12. The number of non-trivial components was determined by comparing eigenvalues to the maximum eigenvalue of a PCA using randomized beta values (shuffling of sample labels per probe) (Extended Data Fig. 1b). Principal component scores for all non-trivial components ($n = 94$) were used for *t*-SNE analysis¹⁷ (*Rtsne* package v.0.11, Fig. 1b). The following non-default parameters were used: $\theta = 0$, $\text{pca} = \text{F}$, $\text{max_iter} = 2500$. A similar approach was used for the combined analysis of reference and diagnostic cases (Fig. 5a).

The random forest algorithm. The random forest¹⁹ algorithm is a so-called ensemble method that combines the predictions of several weak classifiers to achieve improved prediction accuracy. The random forest algorithm uses binary decision trees (classification and regression trees, CART³⁹) as weak classifiers (Extended Data Fig. 4). Each of these trees is a sequence of binary splitting rules that are learned by recursive binary splitting. The CART algorithm starts with all samples assigned to a 'root' node and tries to find the variable, for example, a measured CpG probe, and a corresponding cutoff that results in the purest split into the different classes. To measure this gain in class 'purity', the Gini index is used. To fit a tree, the CART algorithm iteratively repeats these steps until no further improvements can be made. To predict the class of a new diagnostic case, the binary splitting rules are compared with the new data starting in the root node down to one of the leaf nodes. The tree then predicts or votes for the class of that leaf node. Decision trees have the advantage that they are non-parametric and do not rely on any distributional assumptions. The main disadvantages of decision trees are that they often tend to overfit the data and that they have a weak prediction performance. To improve the prediction accuracy, the random forest algorithm combines thousands of trees by bootstrap aggregation (bagging). In brief, each tree is fitted using training datasets that are generated by drawing bootstrap samples. In addition, at each node only a random subset of the available variables is used to find an optimal splitting rule. This additional source of randomization allows selecting variables with lower predictive value. This feature guarantees that the resulting trees are decorrelated, that is, they use different variables to find an optimal prediction rule. Taking the majority vote over thousands of bootstrap-aggregated and decorrelated trees greatly improves the prediction accuracy of the random forest.

Classifier development. To train the random forest classifier, the *randomForest* R package⁴⁰ was used. First, the most important features (probes) were selected by applying the random forest algorithm to the beta values of all filtered 428,799 probes. For efficient computation, the probes were split into 43 sets of approximately 10,000 probes. For each set, 100 trees were fitted using 654 randomly sampled candidate features at each split (*mtry* parameter, square root of 428,799, as would be used by default when not splitting into sets). To take the imbalanced methylation class sizes into account, a down-sampling strategy was followed that ensures an identical number of samples per class (parameter *sampsiz* = *rep*(8, 91)), eight reflecting the minimum number of cases in the 91 classes⁴¹. For all other parameters the default settings were used. This procedure was repeated 100 times, essentially fitting 10,000 trees per probe. Finally, features are selected by the permutation-based variable importance measure as implemented in the *randomForest* R package⁴⁰. The importance measure is the class-specific mean decrease in classification accuracy when the feature is permuted. We selected features by ranking them using the minimal rank of the variable importance measures across all classes.

The final random forest classifier was trained by fitting 10,000 trees with the parameter *mtry* = 100 using beta values of the 10,000 probes selected during feature selection. Imbalanced class sizes were accounted for by down-sampling (as described above), and for all other parameters the default settings were used. An overview of the processes is given in Extended Data Fig. 4.

Classifier cross-validation. Overfitting of the training data is a typical problem expected when training classifiers on high-dimensional data. As it often cannot be avoided, the typical strategy to deal with this problem is to evaluate the model accuracy on an independent test dataset or apply cross-validation methods⁴². Because some of the newly defined methylation groups presented in this work cannot be diagnosed by classical histopathological methods or other established molecular assays, an independent test set to assess model accuracy is not available. Therefore, the accuracy of the presented random forest model with the accompanying calibration model was evaluated by a threefold, nested cross-validation. For this, the reference dataset was split into three equally sized parts. In each cross-validation iteration, two-thirds of the data were used to train a random forest classifier in the same way as the random forest classifier for the complete dataset was trained. Then, the remaining one-third of the data were used for predictions using this random forest classifier. After the third iteration of the cross-validation is completed, each

of the 2,801 reference samples has been predicted by an independent random forest classifier, that is, where the sample was not used for estimating batch effects, performing variable selection, or training of the classifier.

Classifier score calibration. The classification scores generated by our multiclass random forest model (that is, the proportion of trees voting for a class) performed well when they were used to assign the correct class labels, but they do not reflect well calibrated class probabilities⁴³. Furthermore, the distribution of the random forest scores varied between classes, which made inter-class comparisons difficult. Moreover, to evaluate a diagnostic classification, the uncertainties associated with an individual prediction in terms of confidence scores or estimated class probabilities are needed.

To obtain scores that are comparable between classes and that are improved estimates of the certainty of individual predictions, we performed a classification score recalibration by mapping the original scores to more accurate class probabilities^{44,45}. To find such a mapping, an L2-penalized, multinomial, logistic regression model was fitted, which takes the methylation class as response variable and the random forest scores as explanatory variables. The R package *glmnet*⁴⁶ was used to fit this model. In addition, the model was fitted by incorporating a small ridge-penalty (L2) on the likelihood to prevent overfitting, as well as to stabilize estimation in situations in which classes are perfectly separable. The amount of this regularization, that is, the penalization parameter, is determined by running a tenfold cross-validation and choosing the largest value that lies within one standard error of the minimum cross-validation error. Independent random forest scores are needed to fit this model, that is, the scores need to be generated by a random forest classifier that was not trained using the same samples, otherwise the random forest scores will be systematically biased and not comparable to scores of unseen cases. As such, random forest scores generated by the threefold cross-validation are used.

To validate the class predictions generated by using the recalibrated scores of the calibration model, a nested threefold cross-validation loop is incorporated into the main threefold cross-validation that validated the random forest classifier (Extended Data Fig. 4). Within each cross-validation run this nested threefold cross-validation is applied to generate independent random forest scores, which are then used to train a calibration model. The predicted random forest scores resulting from predicting the one-third test data of the outer cross-validation loop are then recalibrated by applying the calibration model that was fitted on the random forest scores generated during the nested cross-validation. A similar cross-validation scheme was used previously⁴⁷ to validate estimated classification probabilities.

Classifier performance measures. Performances of the resulting classifier predictions and scores generated by the cross-validation were assessed by the misclassification error, multiclass area under receiver operating characteristic (ROC) curve (AUC) and the multiclass Brier score. The misclassification error measures the frequency of falsely assigned class labels when using the maximum of the random forest scores or re-calibrated scores as a cutoff to determine the predicted class, that is, the majority vote. To measure the AUC for our multiclass random forest the generalization of the AUC for multiclass classification problems⁴⁸ was used. To measure how well the resulting random forest scores and recalibrated scores perform when used as class probabilities, the multiclass Brier score^{42,43,49} was used. The Brier score is the mean-squared difference between the actual and the predicted class probability and thus measures the same characteristic as the mean squared error measures for a continuous forecast.

Methylation class families. We observed that the majority of misclassification errors occurred within eight groups of histologically and biologically closely related tumour classes. We therefore defined eight methylation class families (MCF). Since calibrated scores represent class probabilities, it is possible to apply the addition rule of well calibrated class probability estimates⁴³ within one MCF to get a class probability for the MCF.

Threshold analysis. Finding an optimal cutoff for diagnostic tests usually involves finding an optimal trade-off between sensitivity and specificity. If there are no preferences regarding specificity or sensitivity, the optimal cutoff is chosen by the upper left corner of the ROC curve or by maximizing the Youden index (specificity + sensitivity - 1). In an application like the one described here, where the cost of false negative is that a tumour cannot be classified and the cost of a false positive is a falsely predicted methylation class, a threshold with high specificity is preferred. ROC analysis is typically defined for binary classification problems. Finding a threshold for multiclass classifiers either involves performing a ROC analysis for each class resulting in class-wise individual thresholds or finding some common threshold for all classes.

The calibrated MC/MCF scores (here referring to MCF and methylation classes (MC) that are not assigned to a MCF) are already validated probability estimates for the methylation class with a direct interpretation, that is, we expect among all samples with scores of approximately 0.9 that 10% are falsely predicted. Applying an additional threshold is not required from a statistical point of view, but is desired in clinical practice. In addition, owing to calibration, scores are comparable across

classes and it is thus reasonable to define a common threshold for all classes instead of finding the optimal cutoff for each individual methylation class.

To determine a common threshold for the calibrated MC/MCF scores, we performed a ROC analysis of the maximum calibrated MC/MCF scores calculated via cross-validation. For this ROC analysis, we defined a new binary class, that is, samples correctly classified during the cross-validation using the maximum calibrated MC/MCF score for classification were considered as 'classifiable' and samples falsely classified using this score were considered 'non-classifiable'.

Following this ROC analysis approach, we determined a cutoff of 0.836 that maximizes the Youden index with a specificity of 93.8% and sensitivity of 93.4% (Extended Data Fig. 5d, e). A maximum specificity of 100% with a sensitivity of 82.7% can be achieved with a threshold of 0.958. Bootstrapped 95% confidence intervals (grey area in Extended Data Fig. 5d) demonstrate the uncertainty of sensitivity and specificity estimates, especially in the upper left corner of the ROC figure, where the considered thresholds are located.

Both thresholds have been determined by cross-validation on our training data of high quality, but real life diagnostic samples were found to achieve slightly lower scores, due to a number of factors that we cannot control, such as lower overall sample quality and lower tumour purity compared to the samples in our reference cohort. Therefore, we decided to lower the maximum specificity threshold to allow a wider spectrum of samples to become a match. For this, we chose a threshold of ≥ 0.9 , which lies in the middle between the Youden index and the threshold for maximum specificity.

Comparison to TCGA pan-glioma methylation classes. To compare our methylation-based classification of CNS tumours with described methylation classes of brain tumours by The Cancer Genome Atlas (TCGA) project, we downloaded the pre-processed methylation dataset that has previously been published¹⁸, including methylation data of 418 low-grade glioma and 377 glioblastoma samples, which had been analysed using the Illumina 450k array or 27k array platforms. To classify our samples according to the TCGA pan-glioma DNA methylation classification, we trained a random forest classifier on this dataset using the 1,300 CpG probe signature provided by the authors and using the default settings of the random forest algorithms implemented in the R package *randomForest*. The results of this classification for astrocytomas, oligodendrogliomas and glioblastomas are shown in Extended Data Fig. 3d and are provided on a case-by-case basis in Supplementary Tables 2, 4.

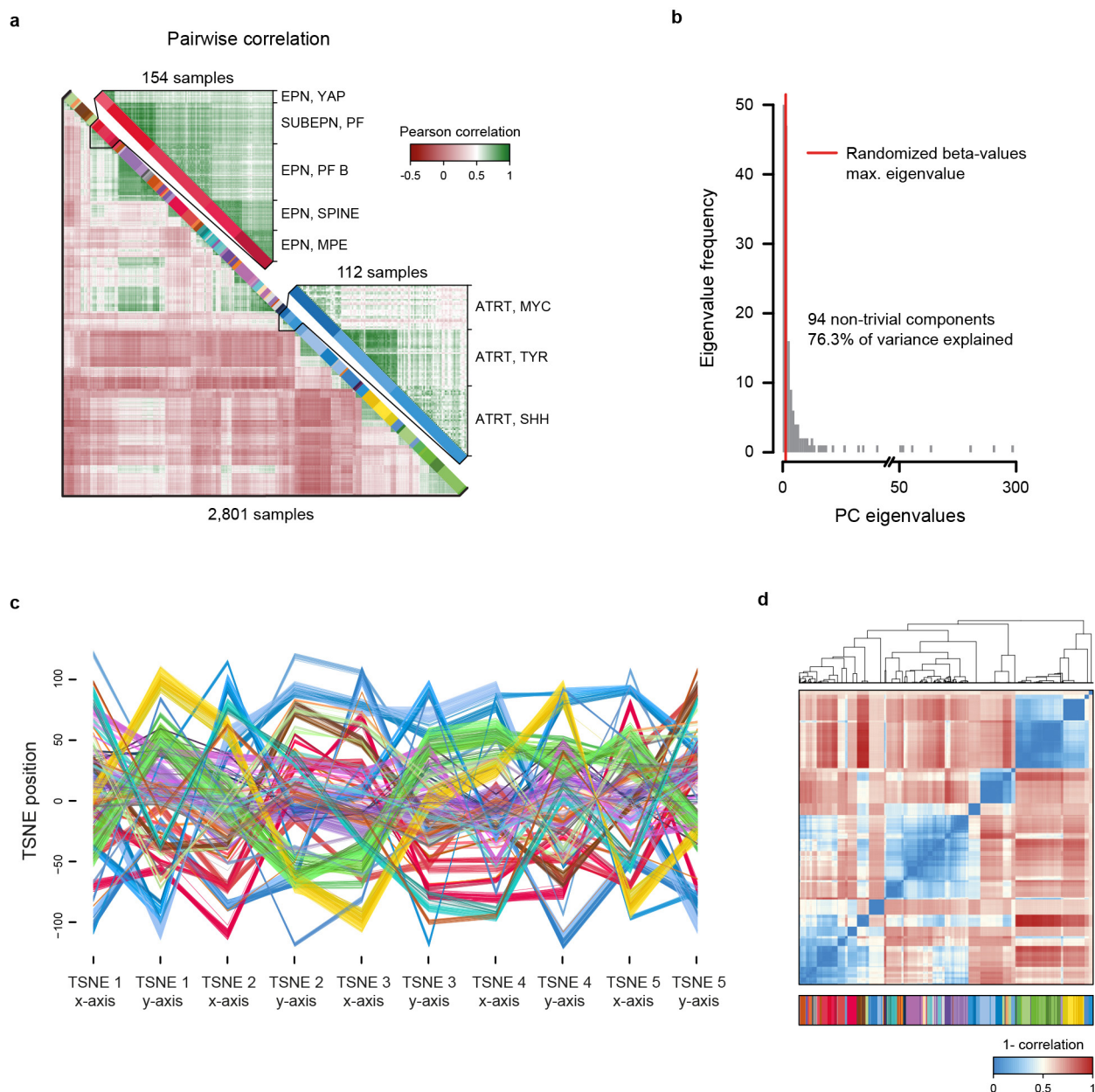
Estimating tumour purity from DNA methylation data. Because of the subjective nature of histological assessment of tumour purity, we additionally used the previously published dataset¹⁸ to train a random forest regression (continuous response variable) model to predict tumour purity⁵¹. This random forest was trained on the 1,000 most important CpG probes for purity estimation selected also by a random forest (similar to the variable selection described for the random forest classifier). The out-of-bag (that is, random forest trees in which the respective sample, for which purity is predicted, was not used for training) mean squared error of the final model was 0.015, indicating that this model was able to yield reasonable predictions of tumour purity from methylation data (Extended Data Fig. 3a–c). The estimated tumour purities for individual cases are given in Supplementary Tables 2, 4.

Code availability. The generated code is available from the corresponding authors upon reasonable request for non-commercial use.

Data availability. The complete methylation values required for the construction of the classifier (reference set) as well as the prospective cohort (validation set) have been deposited in GEO (GSE109381). Supplementary Tables 2 (reference cohort) and 4 (prospective validation cohort) include the IDAT-file names for assignment to patient characteristics. Source Data for Figs 1b, 2b, 3b, 4, 5a, c and Extended Data Figs 1c, 2a–f, 3a–d, 5a, b, d, e, 6 and 7a, c, d are provided with the online version of the paper.

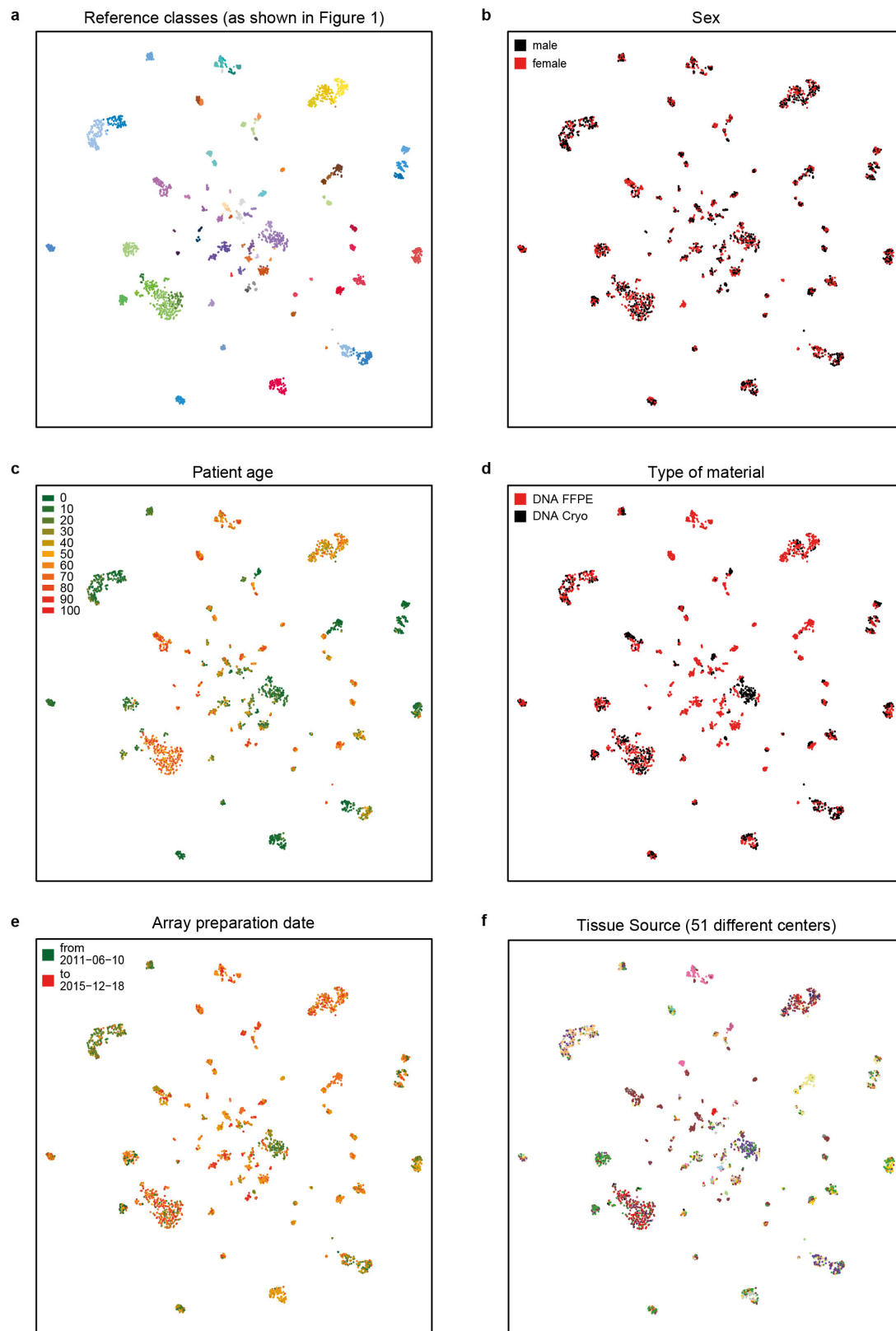
26. Korshunov, A. *et al.* Histologically distinct neuroepithelial tumors with histone 3 G34 mutation are molecularly similar and comprise a single nosologic entity. *Acta Neuropathol.* **131**, 137–146 (2016).
27. Korshunov, A. *et al.* Embryonal tumor with abundant neuropil and true rosettes (ETANTR), ependymoblastoma, and medulloepithelioma share molecular similarity and comprise a single clinicopathological entity. *Acta Neuropathol.* **128**, 279–289 (2014).
28. Höltsken, A. *et al.* Adamantinomatous and papillary craniopharyngiomas are characterized by distinct epigenomic as well as mutational and transcriptomic profiles. *Acta Neuropathol. Commun.* **4**, 20 (2016).
29. Heim, S. *et al.* Papillary Tumor of the pineal region: a distinct molecular entity. *Brain Pathol.* **26**, 199–205 (2016).
30. Koelsche, C. *et al.* Melanotic tumors of the nervous system are characterized by distinct mutational, chromosomal and epigenomic profiles. *Brain Pathol.* **25**, 202–208 (2015).
31. Jones, D. T. *et al.* Recurrent somatic alterations of *FGFR1* and *NTRK2* in pilocytic astrocytoma. *Nat. Genet.* **45**, 927–932 (2013).
32. Jones, D. T. *et al.* Dissecting the genomic complexity underlying medulloblastoma. *Nature* **488**, 100–105 (2012).

33. Pietsch, T. *et al.* Prognostic significance of clinical, histopathological, and molecular characteristics of medulloblastomas in the prospective HIT2000 multicenter clinical trial cohort. *Acta Neuropathol.* **128**, 137–149 (2014).
34. R Core Team. R: A Language and Environment for Statistical Computing. <http://www.R-project.org/> (R Foundation for Statistical Computing, Vienna, Austria, 2016).
35. Huber, W. *et al.* Orchestrating high-throughput genomic analysis with Bioconductor. *Nat. Methods* **12**, 115–121 (2015).
36. Aryee, M. J. *et al.* Minfi: a flexible and comprehensive Bioconductor package for the analysis of Infinium DNA methylation microarrays. *Bioinformatics* **30**, 1363–1369 (2014).
37. Leek, J. T. & Storey, J. D. Capturing heterogeneity in gene expression studies by surrogate variable analysis. *PLoS Genet.* **3**, e161 (2007).
38. Leek, J. T. & Storey, J. D. A general framework for multiple testing dependence. *Proc. Natl Acad. Sci. USA* **105**, 18718–18723 (2008).
39. Breiman, L. *Classification and Regression Trees* (Chapman & Hall/CRC, 1984).
40. Liaw, A. & Wiener, M. Classification and Regression by randomForest. *R News* **2/3**, 18–22 (2002).
41. Chen, C., Liaw, A. & Breiman, L. *Using Random Forest to Learn Imbalanced Data*. Report 666 (Univ. California, Berkeley, 2004).
42. Kim, K. I. & Simon, R. Overfitting, generalization, and MSE in class probability estimation with high-dimensional data. *Biom. J.* **56**, 256–269 (2014).
43. Simon, R. Class probability estimation for medical studies. *Biom. J.* **56**, 597–600 (2014).
44. Boström, H. Calibrating Random Forests. In *Proc. 7th International Conference on Machine Learning and Applications* 121–126 (ICMLA, 2008).
45. Smola, A. J. *Advances in Large Margin Classifiers* (MIT press, 2000).
46. Friedman, J., Hastie, T. & Tibshirani, R. Regularization paths for generalized linear models via coordinate descent. *J. Stat. Softw.* **33**, 1–22 (2010).
47. Appel, I. J., Gronwald, W. & Spang, R. Estimating classification probabilities in high-dimensional diagnostic studies. *Bioinformatics* **27**, 2563–2570 (2011).
48. Hand, D. J. & Till, R. J. A simple generalisation of the area under the ROC curve for multiple class classification problems. *Mach. Learn.* **45**, 171–186 (2001).
49. Brier, G. W. Verification of forecasts expressed in terms of probability. *Mon. Weath. Rev.* **78**, 1–3 (1950).
50. Carter, S. L. *et al.* Absolute quantification of somatic DNA alterations in human cancer. *Nat. Biotechnol.* **30**, 413–421 (2012).



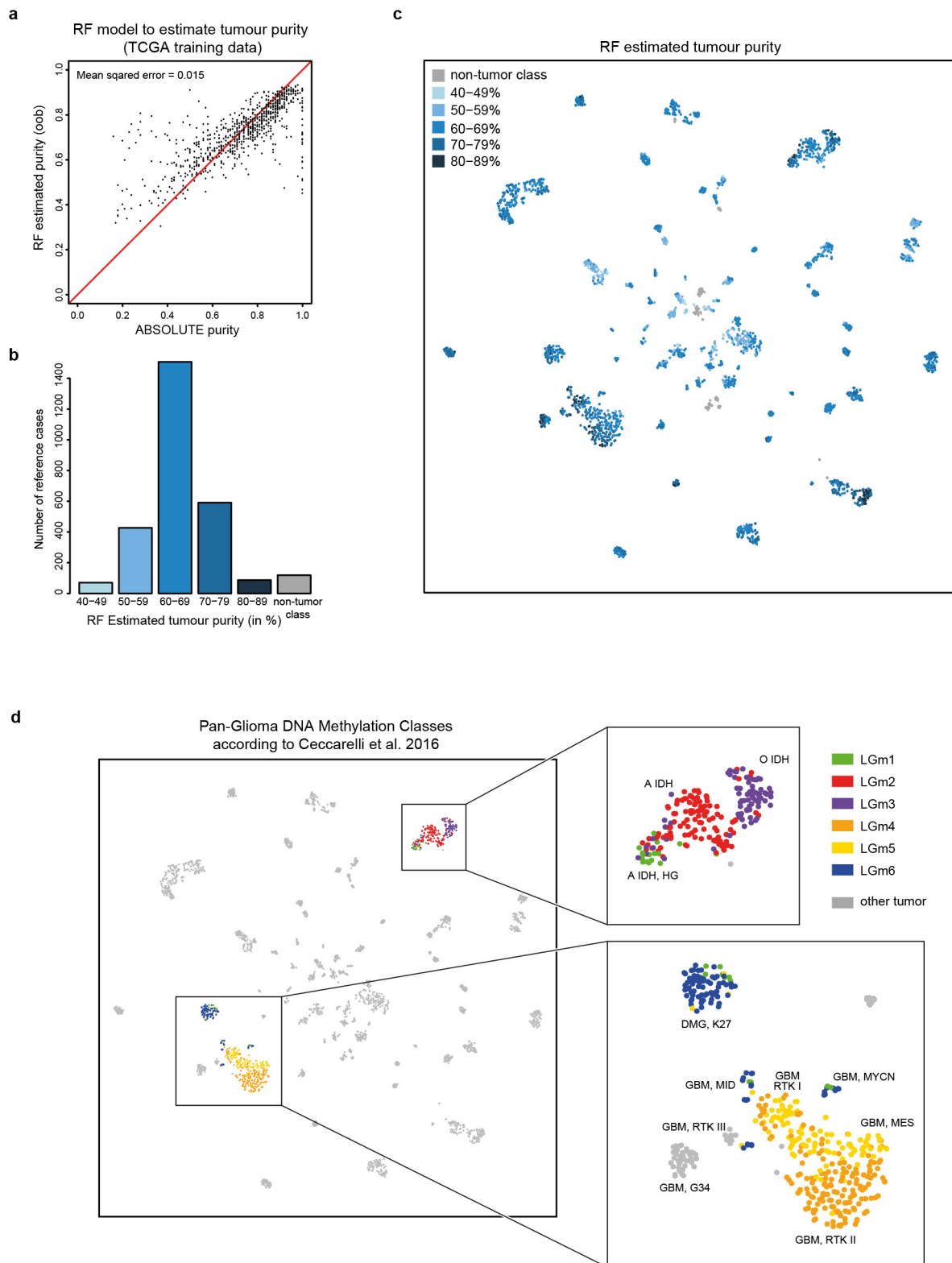
Extended Data Figure 1 | Unsupervised clustering of the DNA methylation-based reference cohort. **a**, Heat map showing the pairwise Pearson correlation (bottom left) of the 32,000 most variably methylated CpG probes of all 2,801 biologically independent samples of the reference cohort. A detailed view of closely related ependymal classes (top right) and the three subclasses identified in atypical teratoid rhabdoid tumours (ATRTs) (bottom right) indicates higher correlation within classes. The colour code and abbreviations are identical to Fig. 1a. **b**, Eigenvalue frequencies of a PCA using the 32,000 most variably methylated CpG probes of all 2,801 biologically independent samples as in **a**. The number of non-trivial components was determined by comparing eigenvalues to the maximum eigenvalue of a PCA using randomized beta values

(shuffling of sample labels per probe). **c**, x and y coordinates of the first five of a total of 500 iterations of t -SNE dimensionality reduction generated by random down-sampling to 90% of the 2,801 biologically independent samples to assess clustering stability. Axis positions of individual cases are connected by a line coloured according to the colour code of Fig. 1a. The depiction illustrates the close proximity of cases of the same class across iterations, indicative of a high stability independent of the exact composition of the reference cohort. **d**, Pairwise correlation of x and y coordinates between 2,801 biologically independent samples over all iterations of the down-sampling analysis demonstrates a very high correlation within classes (average correlation 0.982), indicating a high stability of the t -SNE analysis.



Extended Data Figure 2 | Unsupervised clustering is not biased by a range of possible confounding factors. **a**, *t*-SNE representations of the 2,801 biologically independent samples constituting the reference cohort as shown in Fig. 1b overlaid with potentially confounding factors (**b**–**f**). **b**, Distribution of patient sex among the classes illustrates equal or near equal distribution of many classes, but also an expected enrichment for one sex in some classes (for example, female in meningioma or CNS

high-grade neuroepithelial tumours with MN1 alteration). **c**, Patient age illustrates the expected age distribution of many tumour classes. **d**–**f**, The slightly uneven distribution of type of material (for example, pilocytic astrocytoma or meningioma) (**d**), array preparation date (**e**) and tissue source (**f**) are related to the specifics of assembling the reference cohort and do not indicate an apparent confounding effect on the unsupervised clustering.

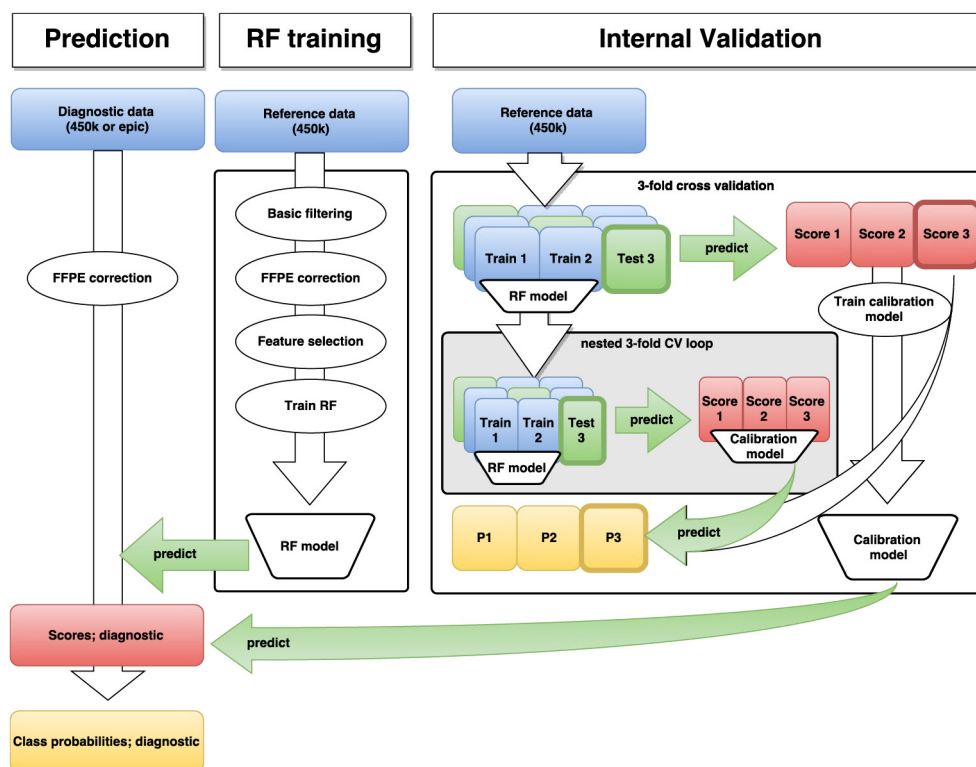


Extended Data Figure 3 | See next page for caption.

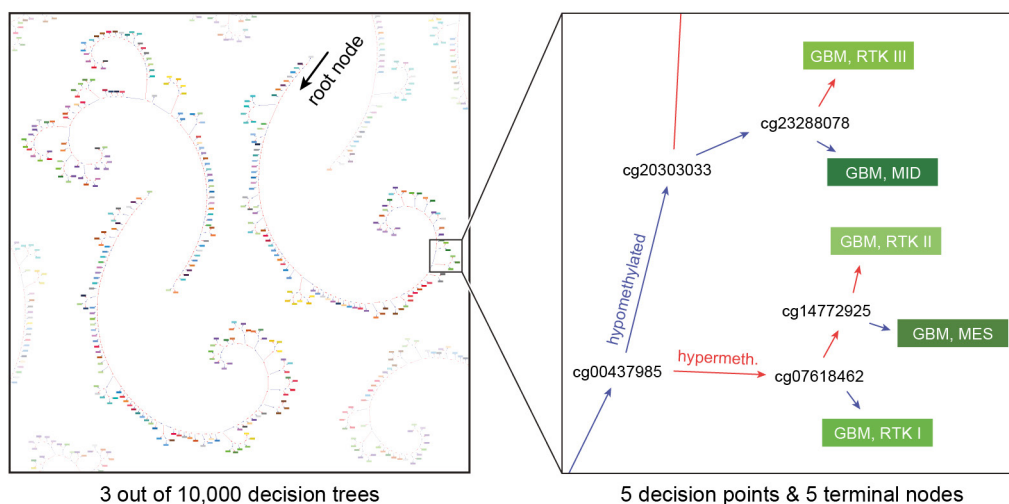
Extended Data Figure 3 | Estimation of tumour purity and relation to TCGA pan-glioma methylation classes. **a**, A random forest model was trained to predict absolute tumour purity estimates⁵⁰ using the TCGA pan-glioma dataset (795 biologically independent samples)¹⁸. The plot shows absolute purity estimates and out-of-bag random forest tumour purity predictions (that is, using only random forest trees for which the respective sample was not involved in the training). The estimated mean squared error is 0.015, indicating that this model is able to yield reasonable predictions of tumour purity from methylation data. **b**, The distribution of random forest predicted purity in the reference dataset (2,801 biologically independent samples). Purity estimates have been transformed into five categories indicated by different shades of blue. The exact case-by-case values are provided in Supplementary Table 2. The median estimated purity in the reference cohort is 66% (range 42% to 87%) and 78% of samples have an estimated purity of at least 60%. **c**, *t*-SNE representation of the reference cohort (2,801 biologically independent samples) overlaid with random forest predicted purity categories. Methylation classes are generally composed of mixed tumour purity categories. Tumour purity shows some association with the WHO grade (WHO I median tumour purity 60%, range 39–77%; WHO II median 66%, range 43–80%; WHO

III median 68%, range 54–84%; WHO IV median 69%, range 49–87%). A further association of tumour purity with the composition of classes in the unsupervised *t*-SNE analysis was not evident. **d**, *t*-SNE representation of the reference cohort (2,801 biologically independent samples) overlaid with predicted TCGA pan-glioma DNA methylation classes according to the previously published dataset¹⁸. Pan-glioma methylation classes were predicted by training a random forest on the previously published dataset¹⁸, which included methylation data of 418 low-grade glioma and 377 glioblastoma samples that were acquired using the Illumina 450k and 27k platforms. The random forest algorithm was trained using the 1,300 CpG signature as described in ref. 18 and using the default settings of the random forest algorithm implemented in the R package randomForest. Pan-glioma class prediction was only performed for subsets of mostly adult astrocytomas, oligodendrogliomas and glioblastomas (magnified areas) included in the previously published dataset¹⁸. LGm1, LGm2 and LGm3 show a high overlap with the methylation classes A IDH HG, A IDH and O IDH, respectively. LGm4 shows the highest overlap with methylation class GBM RTK II. LGm5 shows the highest overlap with methylation classes GBM MES and GBM RTK I. LGm6 show the highest overlap with DMG K27, GBM MID and GBM MYCN.

a



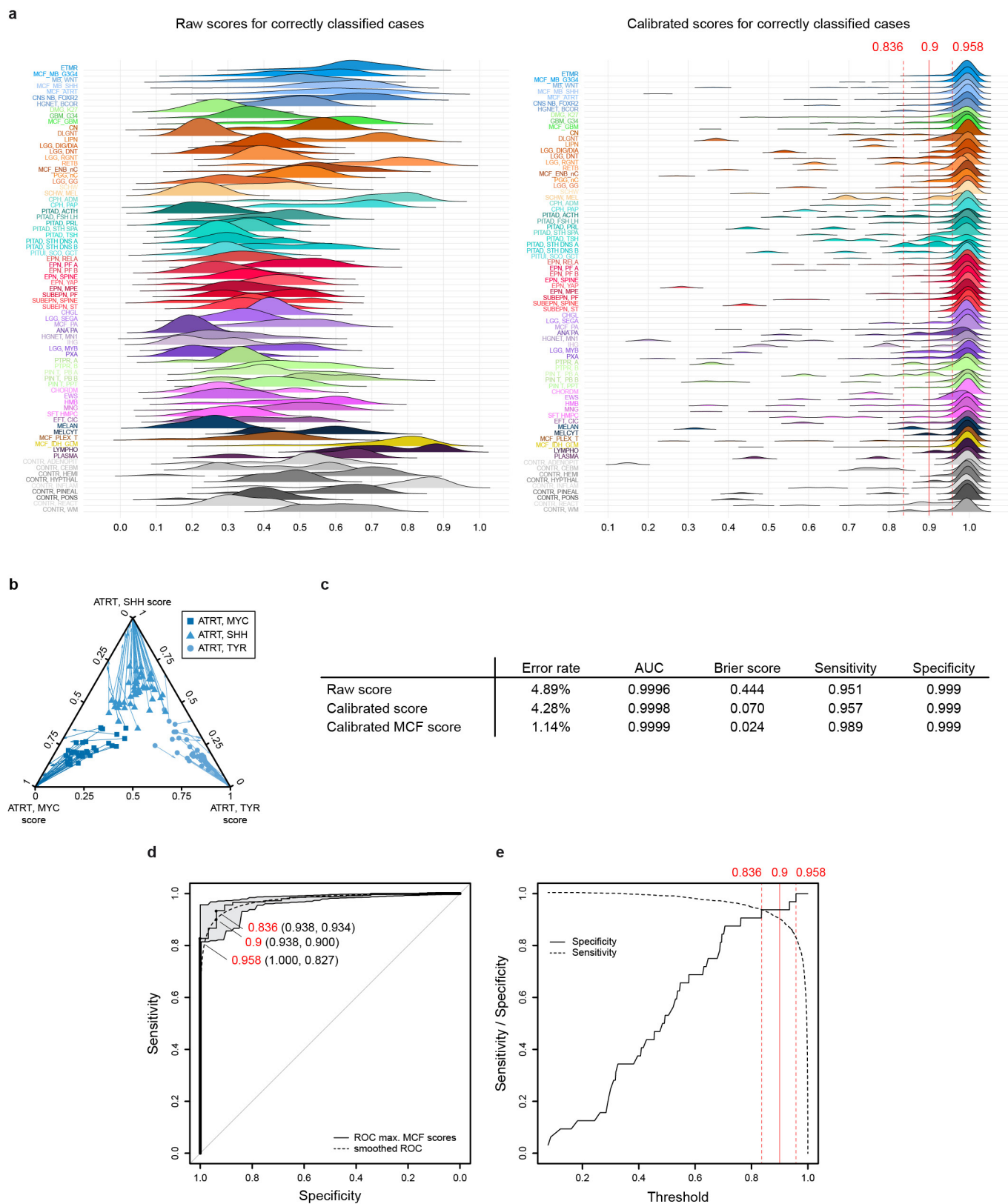
b



Extended Data Figure 4 | Development of the random forest classifier.

a, The random forest training consists of four steps. First, basic filtering of probes that were not included on the EPIC array, probes located on the X and Y chromosomes, probes affected by single nucleotide polymorphisms, and probes not mapping uniquely to the genome was performed. In the second step, the probe-wise batch effects between samples from FFPE and frozen material were estimated and adjusted by a linear model approach. In the third step, feature selection was performed by training a random forest algorithm using all probes and selecting the 10,000 probes with highest variable importance measure. In the last step, the final random forest is trained using only the 10,000 selected probes. The validation of the random forest classifier involves a threefold nested cross-validation. In the outer loop of the cross-validation, the complete random forest training procedure consisting of four steps as described above are applied to the training data and the resulting random forest is used to predict the test data to generate random forest scores. In the inner loop of the cross-validation a threefold cross-validation is applied to

training data of the outer loop in order to generate random forest scores independent of the test data in the outer loop. These scores are then used to fit a calibration model, that is, a L2-penalized, multinomial, logistic regression that takes the random forest scores of the test data in the outer cross-validation loop to estimate tumour class probabilities (P1, P2, P3). To fit a calibration model to estimate class probabilities of diagnostic samples using all data in the reference set, the random forest scores generated in the outer cross-validation loop were used. **b**, Schematic depiction of three example binary decision trees of the random forest classifier (left), and magnification on five example decisions nodes relevant for glioblastoma classification (right). For prediction, a diagnostic sample enters the root node of each of the 10,000 trees. At every decision node, the decision path is determined on the methylation level of a single CpG, until it reaches a terminal node that provides the class prediction. The joint class prediction of all trees represents the raw prediction score. The colour code and abbreviations are identical to Fig. 1a.

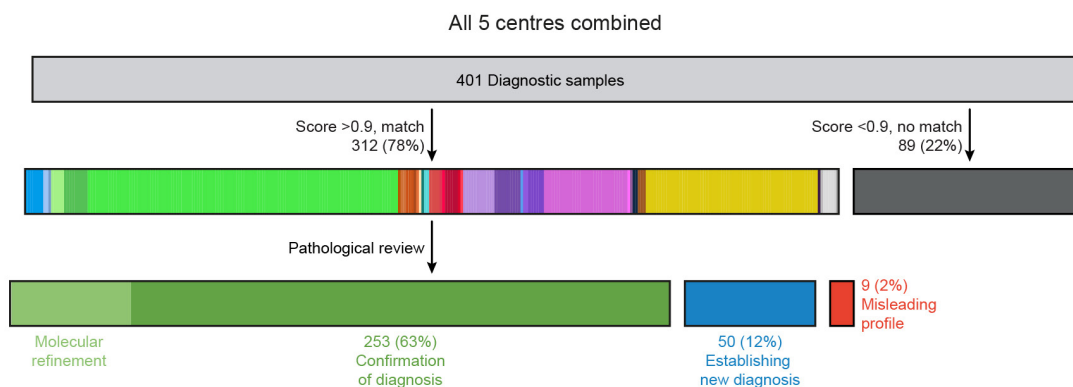


Extended Data Figure 5 | See next page for caption.

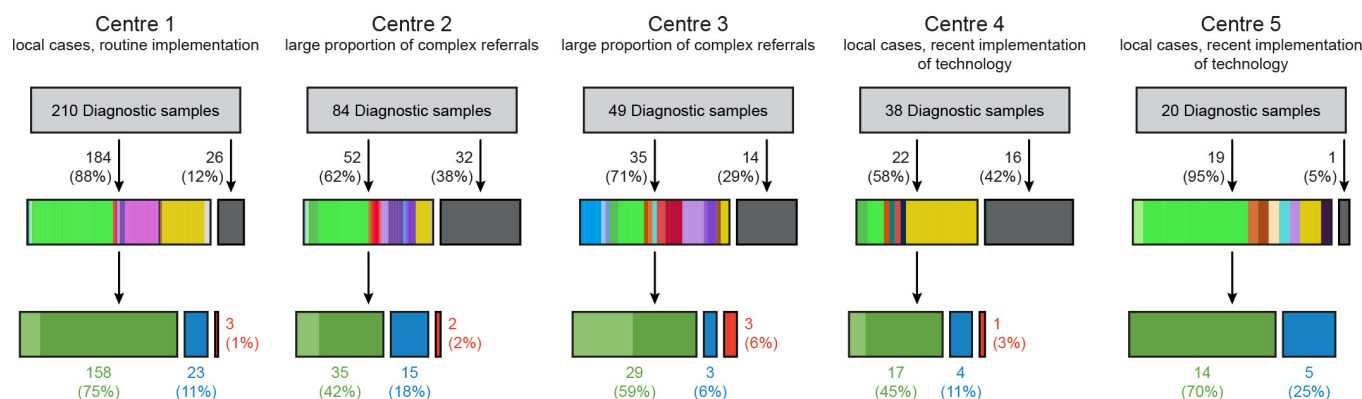
Extended Data Figure 5 | Comparison of raw and calibrated classifier scores and threshold definition. **a**, Density plots illustrating the distribution of raw and calibrated classifier scores for samples correctly classified during cross-validation ($n = 2,701$ independent biological samples for raw and $n = 2,769$ independent biological samples for calibrated), depicted for each methylation class or methylation class family. Score calibration results in a harmonization of score distribution and allows the establishment of a shared classification threshold. Three thresholds for maximizing specificity (0.958), maximizing the Youden index (0.836), and the cutoff used in this study (0.9) are indicated by red lines (see also **d** and **e**). **b**, Multivariate score calibration illustrated as a ternary plot showing scores of the three ATRT subclasses (MYC, SHH and TYR; together $n = 112$ independent biological samples). Arrows indicate transformation of the scores for individual samples by the calibration model, which increases the discrimination between the three subclasses. **c**, The accuracy of prediction of the random forest classifier constructed of $n = 2,801$ biologically independent samples (measured by misclassification error, AUC, Brier score, multiclass sensitivity and specificity) is improved by score calibration and by combining classes into MCF). **d**, To determine

a common threshold for the calibrated MCF scores, we performed a ROC analysis of the maximum calibrated MCF scores of all $n = 2,801$ biologically independent samples calculated via cross-validation. For this ROC analysis, we defined a new binary class, that is, samples correctly classified during the cross-validation using the maximum calibrated MCF score for classification were considered as 'classifiable' ($n = 2,769$) and samples that got falsely classified using this score were considered 'non-classifiable' ($n = 32$). Three thresholds for different sensitivity and specificity are highlighted in the ROC curve: a threshold of 0.958 achieving a maximum specificity of 1 with a sensitivity of 0.827, a threshold of 0.836 obtaining a maximum Youden index with specificity 0.938 and sensitivity 0.934, and our recommended threshold of 0.9 that results in a specificity of 0.938 and a sensitivity of 0.9. Bootstrapped 95% confidence intervals for estimated sensitivity and specificity are indicated in grey. **e**, Sensitivity and specificity for all possible thresholds applied to cross-validated maximum MCF classifier scores of all $n = 2,801$ biologically independent samples. Three thresholds for maximizing specificity (0.958), maximizing the Youden index (0.836) and 0.9 are highlighted by red lines.

a

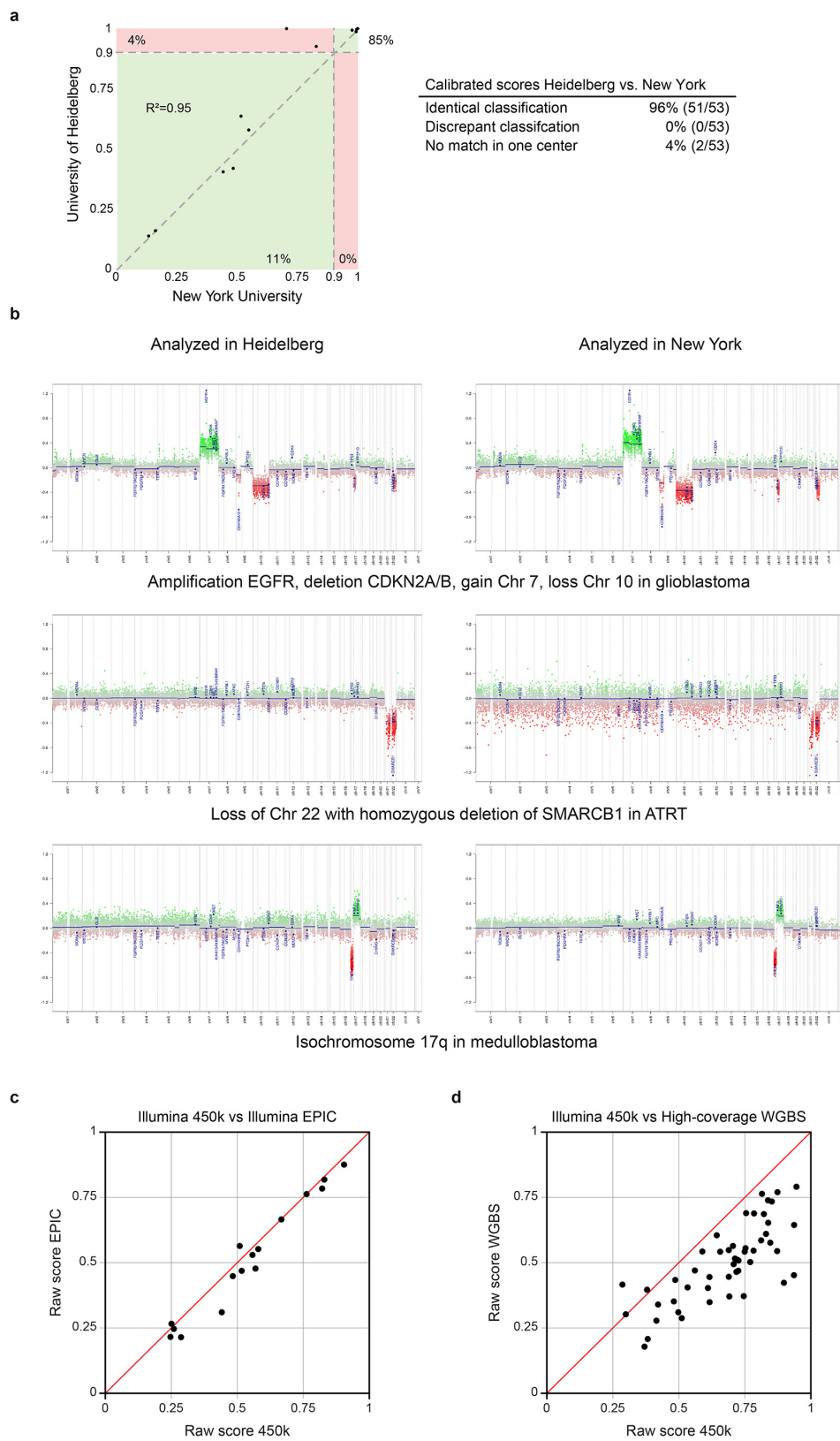


b



Extended Data Figure 6 | Diagnostic utility of the DNA methylation-based classifier, assessed at different centres. a, Implementation of the DNA methylation classifier by five external centres. In total, 401 independent biological samples were analysed. 78% matched to an established class with a cut-off score of ≥ 0.9 (class colours as in Fig. 1a).

A new diagnosis was established in 12% of cases. **b,** Depiction of individual centre results, illustrating the different composition of samples included in the analysis, variation in the rate of non-matching cases, and of cases for which a new diagnosis was established. Case-by-case details are provided in Supplementary Table 6.



Extended Data Figure 7 | See next page for caption.

Extended Data Figure 7 | Inter-centre and inter-platform reproducibility of the DNA methylation-based classification.

a, Calibrated scores of 53 independent biological samples representing diagnostic CNS tumour cases analysed at the University of Heidelberg and at the New York University pathology department. Both laboratories performed independent DNA extraction, array hybridization and data analysis. Cases falling into green areas were classified identically in both centres (96%); cases in the red area were non-classifiable in one centre (4%). None of the 53 samples was assigned to a different methylation class by the two centres. **b**, Copy-number profiles calculated from the array data generated at both centres were highly comparable and allowed identification of chromosomal gains, losses, amplifications and deletions.

Calculations and interpretation were performed once at each centre.

c, Plot of maximum raw classification scores of 16 different tumour samples generated using both 450k and EPIC arrays. All cases fall close to the bisecting line (red) indicating a high concordance of the scores. Further, the methylation class prediction was identical for all samples. **d**, The CNS tumour classifier also performs well with data generated by WGBS. The plot shows classifier scores calculated from WGBS and 450k arrays of 50 cases comprising 11 different brain tumour entities (bisecting line in red). Methylation beta values were calculated from high-coverage WGBS data (>10 fold average coverage) and run through the CNS tumour classifier and plotted against the same case analysed using 450k arrays. The highest class prediction score was identical in all cases.

| Timeline working days (total days) day of week | Process | Comments |
|------------------------------------------------------|-----------------------------------------------------------------------------------------------------|------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| Day 1 (1) Tuesday | Material assessment | Methylation profiling requires high tumour content (>70% tumor cells of total cells) Hematoxylin and eosin staining for identification of optimal area. Scraping of tumour cell rich areas from unstained slides (10 slides of 10 µm each) or punch extraction from paraffin block (either 1.5 mm or 3 mm diameter). Cases with low tumor content are avoided when possible. |
| Day 2 (2) Wednesday | DNA extraction (1/2) | Day one of automated DNA extraction (Maxwell 16 FFPE plus LEV DNA purification kit, AS1135, Promega). |
| Day 3 (3) Thursday | DNA extraction (2/2) DNA quality control Bisulfite conversion (1/2) | Day two of DNA extraction. Concentration measurement using Qubit (Qubit dsDNA BR assay kit, Q32853, Invitrogen). Illumina FFPE QC kit only used optionally (e.g. for highly necrotic cases). Day one of bisulfite conversion of 250ng (in exceptional cases as little as 50-100ng) (Zymo EZ DNA methylation kit; D5002, Zymo). |
| Day 4 (4) Friday | Bisulfite conversion (2/2) DNA Restoration | Day two of bisulfite conversion. FFPE DNA Restore (Illumina). DNA cleanup. Part of methylation array kit provided by Illumina. Performed strictly according to instruction manual. Freezing over weekend. |
| Day 5 (7) Monday | Whole genome amplification | Part of methylation array kit provided by Illumina. Performed strictly according to instruction manual. |
| Day 6 (8) Tuesday | Fragmentation, precipitation Resuspension Array hybridization (1/2) | Part of methylation array kit provided by Illumina. Performed strictly according to instruction manual. |
| Day 7 (9) Wednesday | Array hybridization (2/2) Washing and staining Scanning of array Technical quality control | Part of methylation array kit provided by Illumina. Performed strictly according to instruction manual. Technical quality control using assay internal system controls following the instruction manual. Eliminate bad samples. |
| Day 8 (10) Thursday | Data upload to Webportal Integration with pathological findings Writing of reports | Upload to www.molecularneuropathology.org Gender check. Checking for genotype matches if established. Exclusion of case swapping. Integration of pathological findings with methylation profiling. Possibly reconsideration of histology and possibly additional molecular investigations required. |

Extended Data Figure 9 | Example work flow and timeline of diagnostic methylation profiling.

Integrative structure and functional anatomy of a nuclear pore complex

Seung Joong Kim^{1*}, Javier Fernandez-Martinez^{2*}, Ilona Nudelman^{2*}, Yi Shi^{3†*}, Wenzhu Zhang^{3*}, Barak Raveh¹, Thurston Herricks⁴, Brian D. Slaughter⁵, Joanna A. Hogan⁶, Paula Upla⁷, Ilan E. Chemmama¹, Riccardo Pellarin^{1†}, Ignacia Echeverria¹, Manjunatha Shivaraju⁵, Azraa S. Chaudhury², Junjie Wang³, Rosemary Williams², Jay R. Unruh⁵, Charles H. Greenberg¹, Erica Y. Jacobs³, Zhiheng Yu⁸, M. Jason de la Cruz^{8†}, Roxana Mironska², David L. Stokes⁷, John D. Aitchison^{4,9}, Martin F. Jarrold⁶, Jennifer L. Gerton⁵, Steven J. Ludtke¹⁰, Christopher W. Akey¹¹, Brian T. Chait³, Andrej Sali¹ & Michael P. Rout²

Nuclear pore complexes play central roles as gatekeepers of RNA and protein transport between the cytoplasm and nucleoplasm. However, their large size and dynamic nature have impeded a full structural and functional elucidation. Here we determined the structure of the entire 552-protein nuclear pore complex of the yeast *Saccharomyces cerevisiae* at sub-nanometre precision by satisfying a wide range of data relating to the molecular arrangement of its constituents. The nuclear pore complex incorporates sturdy diagonal columns and connector cables attached to these columns, imbuing the structure with strength and flexibility. These cables also tie together all other elements of the nuclear pore complex, including membrane-interacting regions, outer rings and RNA-processing platforms. Inwardly directed anchors create a high density of transport factor-docking Phe-Gly repeats in the central channel, organized into distinct functional units. This integrative structure enables us to rationalize the architecture, transport mechanism and evolutionary origins of the nuclear pore complex.

Nuclear pore complexes (NPCs) are large proteinaceous assemblies studded through the nuclear envelope, the double-membraned barrier that surrounds the nucleus; NPCs are the sole mediators of macromolecular transport between the nucleus and the cytoplasm, and carry key regulatory platforms for numerous nuclear processes¹. NPCs are also major targets for viral manipulation and defects in this transport machine are directly linked to human diseases, including cancers². Each NPC is an eight-fold symmetric, cylindrical assembly consisting of approximately 550 copies of about 30 different proteins of the nucleoporin family (Nups). These Nups assemble into sub-complexes that form higher-order structures called spokes. Eight spokes assemble into even larger modules: coaxial outer and inner rings form a symmetric core scaffold, which is connected to a membrane ring, a nuclear basket and cytoplasmic RNA export complexes³. The scaffold surrounds a central channel that is formed in part by multiple intrinsically disordered Phe-Gly (FG) repeat motifs that extend from nucleoporins termed FG Nups. These FG motifs mediate selective nucleocytoplasmic transport through specific interactions with nuclear transport factors (NTFs), which carry their cognate macromolecular cargoes⁴. It has also previously been suggested that the central channel contains a feature called the central transporter⁵. Although partial structures have previously been described^{3,6,7}, a complete, high-resolution structure for the entire NPC in any organism has hitherto been lacking, leaving open key questions as to how the NPC is organized and functions, and how it evolved. To address these questions, we have determined an integrative structure of the yeast NPC at sub-nanometre precision.

Solving the structure of the *S. cerevisiae* NPC

We developed a method to rapidly and gently isolate native yeast NPCs, enabling us to determine the type and amount of each Nup in the NPC, the proximities between Nups resolved to the amino acid residue level, and the mass and detailed morphology of the entire NPC. These data were then used to solve the structure using an integrative modelling approach^{8,9} (Extended Data Fig. 1, Supplementary Results and Discussion, and Methods).

We determined the mass of the entire NPC and a definitive stoichiometry for every Nup and associated molecules using mass spectrometric and *in vivo* imaging methods. The native NPC has a mass of 52 MDa, or about 87 MDa when including the membrane, cargo and NTFs (Fig. 1 and Extended Data Figs 2, 3c). To inform the proximities, orientations and conformations of the Nups, isolated NPCs were subjected to cross-linking with mass spectrometric readout^{9,10}. This approach identified 3,077 unique cross-linked pairs of residues, and provided the distance restraints between them, both within and between Nups (Fig. 2, Supplementary Table 1 and Methods). The morphology of the NPC was determined using cryo-electron tomography (cryo-ET) and sub-tomogram averaging¹¹ (Methods). This approach provided a final 3D map at approximately 28 Å resolution, with a local resolution of 20–25 Å for the inner ring, which has approximate C₂ symmetry (Fig. 3 and Extended Data Figs 4–6). The NPCs retained a considerable amount of nuclear envelope membrane, which forms a continuous belt around the midline of the structure (Fig. 3a, b, d, e). We found that a membrane protein ring interconnects adjacent spokes

¹Department of Bioengineering and Therapeutic Sciences, Department of Pharmaceutical Chemistry, and California Institute for Quantitative Biosciences, University of California, San Francisco, San Francisco, California 94158, USA. ²Laboratory of Cellular and Structural Biology, The Rockefeller University, New York, New York 10065, USA. ³Laboratory of Mass Spectrometry and Gaseous Ion Chemistry, The Rockefeller University, New York, New York 10065, USA. ⁴Institute for Systems Biology, 401 Terry Ave. N., Seattle, Washington 98109, USA. ⁵Stowers Institute for Medical Research, Kansas City, Missouri 64110, USA. ⁶Department of Chemistry, Indiana University, Bloomington, Indiana 47405, USA. ⁷Skirball Institute and Department of Cell Biology, New York University School of Medicine, New York, New York 10016, USA. ⁸Janelia Research Campus, Howard Hughes Medical Institute, 19700 Helix Drive, Ashburn, Virginia 20148, USA. ⁹Center for Infectious Disease Research, Seattle, Washington 98109, USA. ¹⁰Verna and Marrs McLean Department of Biochemistry and Molecular Biology, Baylor College of Medicine, 1 Baylor Plaza, Houston, Texas 77030, USA. ¹¹Department of Physiology and Biophysics, Boston University School of Medicine, 700 Albany Street, Boston, Massachusetts 02118, USA. [†]Present addresses: Structural Bioinformatics Unit, Institut Pasteur, CNRS UMR 3528, Paris, France (R.P.); Department of Cell Biology, University of Pittsburgh School of Medicine, Pittsburgh, Pennsylvania 15260, USA (Y.S.); Structural Biology Program, Sloan Kettering Institute, Memorial Sloan Kettering Cancer Center, New York, New York 10065, USA (M.J.d.l.C.).

*These authors contributed equally to this work.

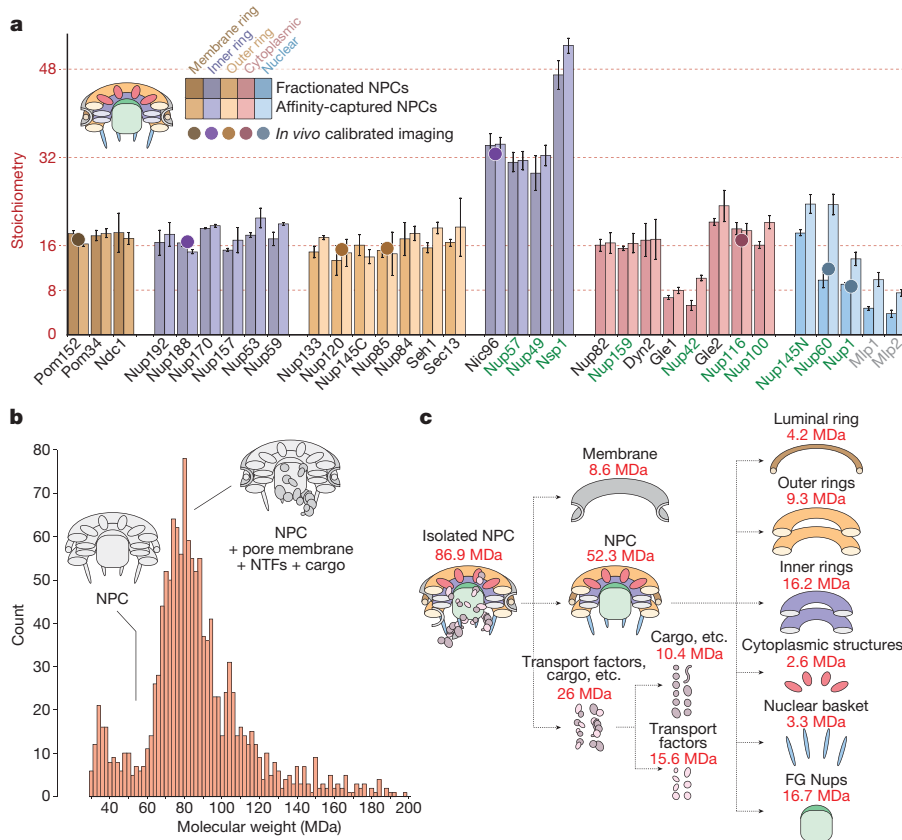


Figure 1 | Defining the mass, composition and stoichiometry of the native NPC.

a, Stoichiometry of the entire complement of NPC components determined by quantitative mass spectrometry (bar plot) and by *in vivo* calibrated imaging of Nup–GFP reporters (dots) (Extended Data Fig. 3a, b). Darker and lighter colour bars (average \pm s.d.) represent measurements from a diploid non-tagged *S. uvarum* strain ($n = 2$ or 3 technical and 2 biological replicas) and haploid tagged *S. cerevisiae* strains ($n = 1$ –3 technical and 4 biological replicas), respectively. Each Nup is coloured on the basis of its localization, as depicted in the cartoon. FG-repeat-containing Nups are labelled in green. **b**, Affinity-captured whole NPCs were analysed intact by charge detection mass spectrometry, and a representative mass spectrum is shown. $n = 2$ biological replicas; more than 3 runs with over 1,500 individual NPCs per run. **c**, Dissection of the mass and composition of an NPC.

within the nuclear envelope lumen (Fig. 3a, e), a feature largely absent from recent electron microscopy maps. A cylindrically averaged, bi-lobed density fills the central channel (the ‘central transporter’, Fig. 3a, b). Individual Nups and their domains, as well as the sub-complexes of the NPC, were represented on the basis of published crystallographic structures, integrative structures and comparative models^{9,10,12} (Supplementary Table 2 and Methods), and validated by small-angle X-ray scattering profiles for 18 Nups (147 constructs; Supplementary Table 6 and Methods). An ensemble of structural solutions for the NPC that sufficiently satisfied all experimental data was calculated by extensive configurational sampling^{8,9} (Supplementary Table 3 and Methods). Variability among these solutions defines the precision of our structure, as quantified by the average root-mean-square deviation between solutions in the final ensemble⁹. Our final structure defines the positions of 552 Nups (Fig. 4 and Supplementary Videos 1–3), with an overall precision of about 9 Å (Extended Data Fig. 1e, f). The centroid solution is used as the representative structure. The structure was validated by numerous independent tests (Extended Data Figs 1, 7, 8, Supplementary Tables 3, 4 and Methods).

The multiple functionalities and enormous size of the NPC present unique and substantial structural challenges: it must form a stable passageway with a fixed inner diameter; it must be anchored to the nuclear envelope and stabilize the pore membrane within which it resides, with a height appropriate for the thickness of the nuclear envelope; it must correctly position the transport machinery; and it must resist stresses that might lead to disassembly or malfunction. Our structure suggests how each of these challenges is met and—by comparison with the vertebrate scaffold⁶—how different organisms may meet these challenges (see the section ‘Evolutionary origin and diversity of the NPC’).

Forming a stable and defined passageway

The fitness defects of strains containing Nup truncations provide an estimate of the structural importance of the truncated regions^{9,13}. We quantified the fitness defect of strains containing systematic truncations of every major symmetric Nup using ODELAY¹⁴ (an automated

phenotypic analysis platform; Extended Data Fig. 9). Results were heat mapped onto the NPC structure to reveal critical elements of NPC stability (Fig. 5a). The inner ring of the NPC contains crucial stabilizing elements, including Nic96, which forms the heart of a diagonally oriented column within each spoke (Fig. 5b) and interacts with every other protein in the inner ring (Fig. 4d–f). This high connectivity explains why Nic96 is an essential keystone, holding in place much of the scaffold of the NPC. The remainder of each diagonal column is made of Nup157 and Nup170, which flank Nic96 (Fig. 5b); Nup157 and Nup170 are functionally redundant, but are synthetically lethal¹⁵ and together form another essential element of the diagonal column. Inter-spoke connections represent a second crucial stabilizing element. Nup192 probably serves as a cross-brace between adjacent spokes (Fig. 5a, c). The N termini of Nup170 and disordered regions of Nup53 and Nup59 also form key connections between adjacent spokes¹⁶ (Figs 4d, 5c). The inter-spoke connections are established largely through small, hinge-like contacts that may confer flexibility to the interface between adjacent spokes; the diagonal arrangement of the central columns may also enable rotation or local flexing (Fig. 5b), by accommodating compression and expansion forces from nuclear envelope distortions and from the central transporter and the transit of cargoes. Nup188 and Nup192 act as radial separators between the Nic96 column and the triple coiled-coil domains of Nsp1, Nup57 and Nup49, which form a discontinuous ring that defines the narrowest part of the passageway and may allow some dilation of the NPC (Fig. 4d–f). This architecture sets a soft upper limit of about 40 nm for the size of cargoes that can transit the NPC⁴.

How the NPC shapes the nuclear envelope

The pore membrane, where the inner and outer membranes of the nuclear envelope join, defines the inner surface of a torus and therefore has both concave and convex curvatures (Fig. 5d). The inner ring is anchored to the pore membrane through membrane-binding motifs (MBMs) on the β -propellers at the N termini of Nup157 and Nup170, and on the C termini of Nup53 and Nup59 (refs. 17–19). These proteins

Circos plot of the 3.077 chemical cross-links in NPC

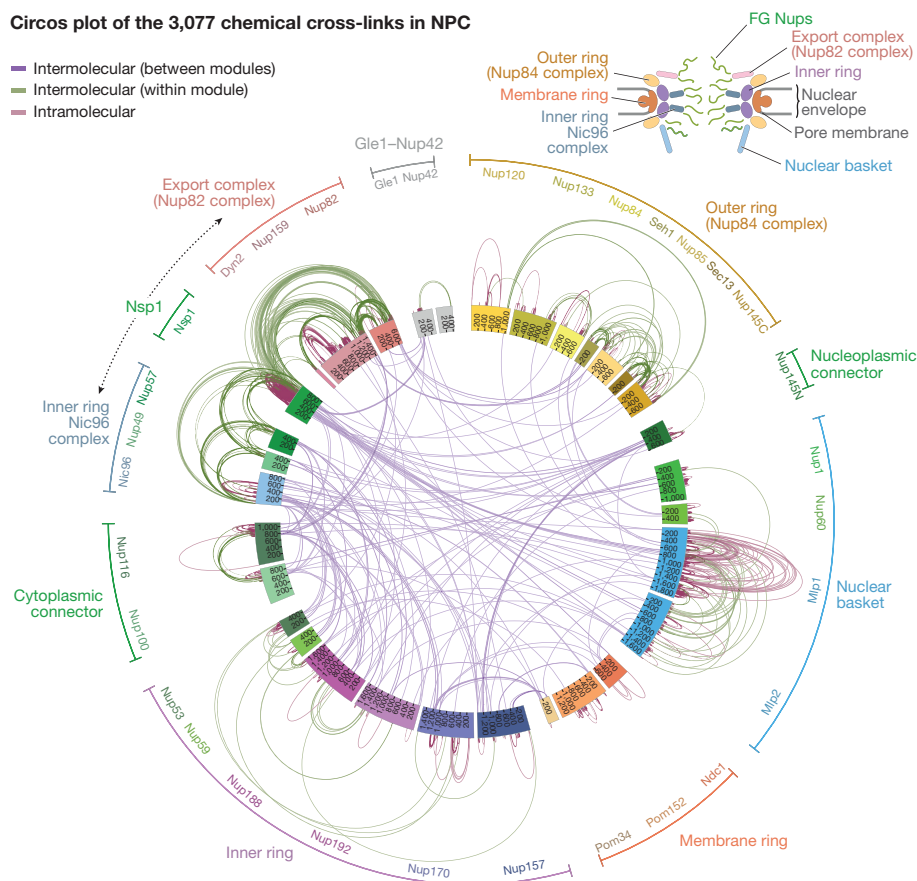


Figure 2 | Chemical cross-linking and mass spectrometry reveals nucleoporin connectivity in the NPC. Circular plot showing the distribution of chemical cross-links (Supplementary Table 1), mapped to each nucleoporin represented as a coloured segment and with the amino acid residues indicated. The identity of each module and Nup is shown in the periphery of the plot. Key in top left gives the types of cross-links; diagram in top right illustrates the relative positions of modules in the NPC.

also interact with the scaffold-facing regions of the transmembrane domain (TMD) carrying Nups, such as Pom152, Ndc1 and Pom34 (Fig. 4d). Together, these MBMs and TMDs form an NPC-anchoring girdle of membrane-associated motifs around the scaffold equator, defining the concave curve of the pore membrane. The convex curvature is defined by both outer and inner rings (Figs 4a, e, 5d). Each outer ring is formed by eight Y-shaped Nup84 complexes arranged head-to-tail and joined by an interaction between the N termini of Nup120 and Nup133 (ref. 20), creating another hinged spoke-to-spoke interface and a minor fitness hotspot (Fig. 5a, c). The outer rings also help define the overall height of the NPC such that it is appropriate for the width of the nuclear envelope. Each Nup84 complex is anchored to the pore membrane by MBMs situated within the N-terminal β -propellers of Nup133 and Nup120^{10,12,21} (Fig. 5d). The convex curvature of the pore membrane is thus defined and stabilized by a ring of MBMs underneath the outer rings and by the thick girdle of MBMs and TMDs around the NPC equator. On the nuclear side, MBMs from Nup1 and Nup60 help anchor the basket to the nuclear envelope²² (Fig. 5d, e).

In the membrane ring, the luminal domain of Pom152 is composed of nine immunoglobulin-like fold repeats²³ that oligomerize in an anti-parallel fashion to form eight circumferential arches within the nuclear envelope lumen, producing additional connections between adjacent spokes. Pom152 appears to be pre-stressed by assembly into these arches (Fig. 4e); the resulting tension may minimize elliptical distortion of the NPC²³. Each arch also delimits a channel (300×120 Å wide) between itself and the underlying pore membrane (Figs 3e, 4e). The outer rings form a series of circumferential arches that align with the Pom152 luminal arches (Fig. 4b, e). These arches align with hinges in the inner ring (Fig. 5c) that could flex to form lateral openings between spokes. This juxtaposition of arches and transient openings may delineate conduits for nucleocytoplasmic transport of transmembrane proteins²⁴, potentially resolving the issue of how membrane proteins transit the NPC²⁵.

Positioning the RNA processing platforms

Whereas the core scaffold is symmetric about the plane of the nuclear envelope, two machineries associated with RNA processing and transport—the basket and export platform—are located at the nuclear and cytoplasmic faces of the NPC, respectively (Fig. 4d–f). At the core of the export platform is the Nup82 complex, which has a coiled-coil bundle that is attached to the Nup85–Seh1 arm and hub region of the Nup84 complex in the cytoplasmic outer ring (Fig. 4f). Together, they form a lateral platform that faces the central channel. An α -helical rod that extends from the Nup82 complex holds Gle1, the RNA helicase Dbp5 and the FG-repeat-carrying Nup42 over the middle of the central channel^{9,25,26}. As a result, numerous transport-factor-docking sites and ATP-dependent RNA remodelling proteins are aligned above the cytoplasmic exit of the NPC to efficiently receive exporting RNAs, to remodel and then release them into the cytoplasm. Likewise, Mlp1 and Mlp2 in the nuclear basket are anchored to the core scaffold mainly by the Nup85–Seh1 arm, similar to the Nup82 complex (Figs 4d, e, 5e). The nuclear basket serves as a platform for the first stages of RNA processing and export²⁷, and the export platform organizes the last stages of export²⁶. Similarities between the export platform and basket suggest that these structures are ancient homologues (Extended Data Fig. 10); their asymmetric localization directs unidirectional export of transcripts out of the nucleus.

Flexible connectors tie the NPC together

Certain disordered connectors have recently been shown to be important for holding parts of the scaffold together^{6,7,16}; the complete structure presented here highlights the extent to which such connectors are critical to NPC integrity. Remarkably, flexible connectors run the entire length of each spoke, tying together every major element in the NPC (Fig. 5c, e). They link the periphery and outer rings to the inner rings, both inner rings to the pore membrane and adjacent spokes to one another. We identified two types of connectors

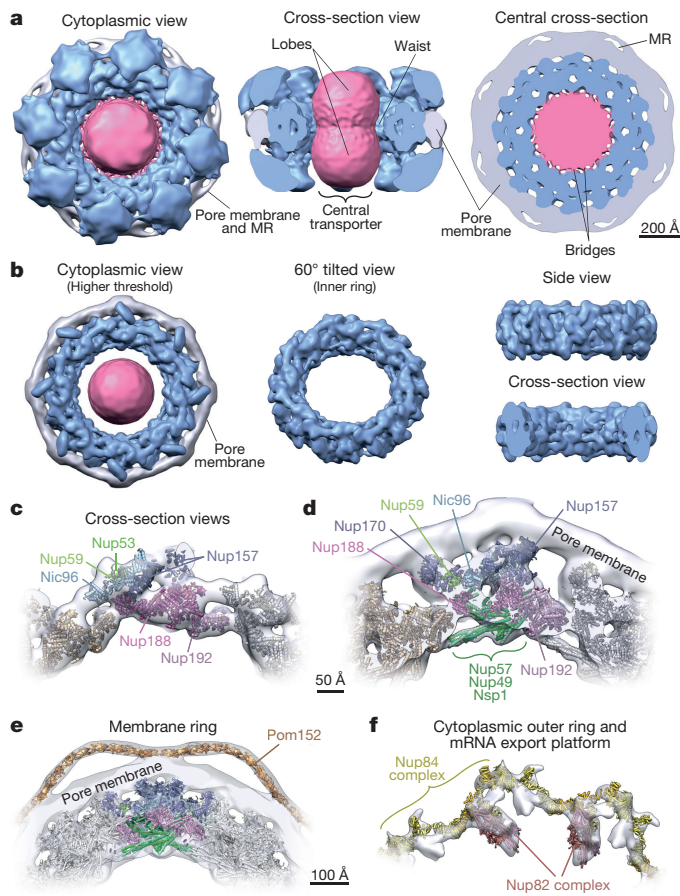


Figure 3 | Morphology of the NPC. **a, b**, Cryo-ET map of the NPC: core scaffold, blue; membrane region, grey; central transporter, pink. MR: membrane ring. In **a**, cytoplasmic top view (left); cross-section side view (middle); and central cross-section top view (right). **b**, Cryo-ET map from **a** presented at a higher threshold. Top view (left); 60°-tilted view of the inner ring (middle); and side (right, top) and cross-section views (right, bottom) of the inner ring. Scale bar, 200 Å. **c–f**, Cross-section views show a representative structure embedded within the cryo-ET density (grey), presented with different filtering and thresholding to show the good fit to the cryo-ET map in the inner ring (**c, d**), the membrane ring (**e**) and the cytoplasmic outer ring and mRNA export platform (**f**). Nups indicated as in Fig. 4. Scale bars, 50 Å (**c, d**) and 100 Å (**e, f**).

(Supplementary Results and Discussion). First, there are vertical connections, aligned parallel to the cylindrical axis of the NPC and constituting the main anchor points between the export platform and the inner ring. On the nuclear side, similar connections are present between the nuclear basket and the inner ring, with an additional connection between the basket and outer ring (Figs 4d, 5e). Second, there are horizontal flexible connectors that link the central channel to the pore membrane between adjacent spokes (Fig. 5e). Collectively, these flexible connectors may serve to allow limited movement of the more rigid modules with respect to one another, thereby providing the NPC with another degree of flexibility in response to deformation²⁸.

Organization of the transport machinery

Despite its critical function, the central gating machinery has largely been excluded from recent NPC maps^{6,7,29} and its properties have remained controversial⁴. Here we confirm the existence of a large central transporter with two high-density ‘lobes’ connected by a narrower ‘waist’ of lower density⁵ (Figs 3, 6 and Extended Data Figs 4–6). This central transporter comprises multiple FG repeats that account for about 9 MDa, together with approximately 26 MDa

of NTFs and their cargoes caught in transit (though they may be somewhat averaged out in our map) (Fig. 1b, c and Extended Data Fig. 3c). Indeed, even after isolation, each NPC carried 10–80 copies of each of the observed NTFs³⁰, reflecting the huge and varied transport flux through NPCs.

The localization of FG-repeat anchor points reveals three patterns. First, a vertical path is formed along each spoke by a continuous array of FG repeats (Fig. 6a, c and Extended Data Fig. 11b–d). By binding to these repeats, NTFs may follow these paths across the entire NPC. Second, the FG anchor points of Nup1–Nup57–Nup49 form a central ring on the equator of the NPC (Fig. 6b). Thin bridges in our cryo-ET map coincide with the location of these FG anchor points, which indicates that these bridges comprise the FG repeats themselves, emanating from their anchor points (Fig. 6b). Third, the structured regions of the NPC largely direct the FG-repeat regions inwards toward the axis of the central channel (Fig. 6a), instead of projecting from the NPC towards the cytoplasm and nucleoplasm as they are often represented⁹. This geometry generates a highly concentrated (25–150 mM) and dynamic FG-repeat phase through which cargo-carrying NTFs readily pass, facilitated by their specific FG interactions, whereas nonspecific macromolecular diffusion is hindered by this same dense phase³¹.

It has previously been suggested that the two main types of FG repeat (‘Phe-X-Phe-Gly/Phe-Gly’ (FXFG/FG) and ‘Gly-Leu-Phe-Gly’ (GLFG)) are segregated in the NPC to define functionally distinct zones of the gating machinery³². Consistent with this, and with the known role of FXFG/FG-type repeats in docking RNAs during export³³, we find that FXFG/FG-type repeats are enriched in the nuclear and cytoplasmic peripheries of the NPC, where the RNA-associating export platform and basket reside (Fig. 6d and Extended Data Fig. 11c). By contrast, the GLFG-type repeats are enriched in regions adjacent to the inner ring and near the cytoplasmic entrance to the central channel. This cytoplasmic localization coincides with the position of FG repeats that are most important for limiting the passage of nonspecific macromolecules (Fig. 6e and Extended Data Fig. 11d), and is consistent with the known role of GLFG-type repeats in maintaining the passive permeability barrier^{34,35}.

Evolutionary origin and diversity of the NPC

NPCs share architectural features with vesicle coating complexes (Extended Data Fig. 10), which led us to hypothesize that they share a common evolutionary ancestor, the ‘protocoatome’³⁶. Two major families of coating complexes exist: COPI/clathrin and COPII, each of which have discrete vesicle recognition and trafficking roles^{37,38}. We find both COPI/clathrin-like and COPII-like features in the NPC, suggesting that ancestral COPI and COPII coating families evolved first and were followed by the NPC, which may have evolved through a partnership of COPI and COPII coats. This hypothesis implies that the nucleus was a later addition in the evolutionary path of the first eukaryotes (Supplementary Results and Discussion).

Despite substantial conservation of some elements of NPC architecture, other elements can vary widely between species. Generally, the inner ring appears most conserved^{39,40}, as is seen in a comparison of our yeast structure with that of the human scaffold⁶, although the latter is more expanded (Extended Data Fig. 12). By contrast, peripheral elements exhibit considerable lineage-specific losses and duplications^{40,41}. In yeast, each outer ring is formed by 8 copies of the Nup84 complex (Figs 3f, 4), whereas in vertebrates each outer ring contains 16 copies of the equivalent Y-shaped complex arranged in two interlocked rings^{17,42}. Moreover, we see neither an additional copy of Nup157 or Nup170 connecting the outer and inner rings nor Nup188 or Nup192 in the outer rings, as indicated in humans⁶ (Fig. 4d–f and Extended Data Fig. 12a). Another previous model assumed that fungal and human core scaffolds have essentially identical structures⁷. Our data invalidate this assumption (Fig. 1, Supplementary Table 2a and Supplementary Results and Discussion), as well as an earlier ‘fencepost’ model⁴³.

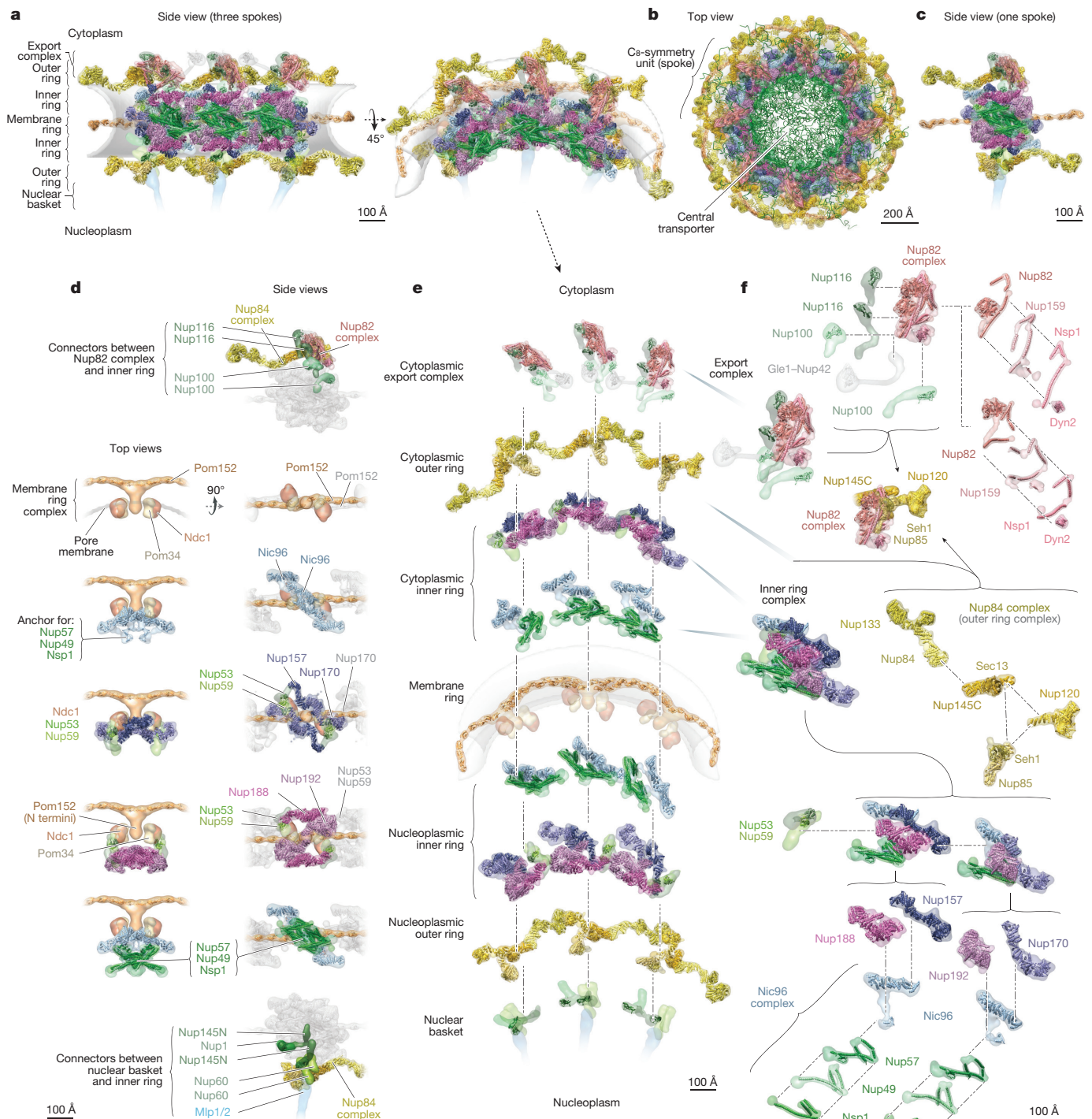


Figure 4 | Structural dissection of the NPC. Complete structure of the NPC and its components shown in different orientations, with a model of the pore membrane region shown in grey (Supplementary Videos 1–3). For each Nup, the localization probability density of the ensemble of structures is shown with a representative structure from the ensemble embedded within it (Supplementary Table 2). **a**, Two views of three consecutive NPC spokes (C_8 -symmetry units), showing how the coaxial outer, inner and membrane rings run continuously between spokes. **b**, Cytoplasmic top view of the complete NPC structure with modelled FG-repeat regions (green). **c**, Side view of a single NPC spoke. **d**, Relative position of major NPC components and connections

both within and between spokes. Top (left) and side (right) views are shown. The membrane ring (beige) is included for reference. Flexible connectors between outer and inner rings are shown in the top and bottom panels, with the inner and membrane rings shown as faded grey densities. **e**, Exploded view of three consecutive spokes, spanning from the cytoplasmic face (top) to the nuclear face (bottom), with dashed lines connecting neighbouring rings. **f**, Cytoplasmic mRNA export complex (top), the Nup84 complex (centre) and the inner ring complex, including the Nic96 complex (bottom), from a single spoke. The complexes are shown as an exploded diagram, with dashed lines connecting neighbouring components.

In summary, there is no single universal NPC structure; instead, similar structural elements are used in somewhat different arrangements to generate many lineage-specific adaptations.

Conclusions

We have described the structure of the entire yeast NPC at sub-nano-metre precision. At the heart of the inner ring, rigid diagonal columns

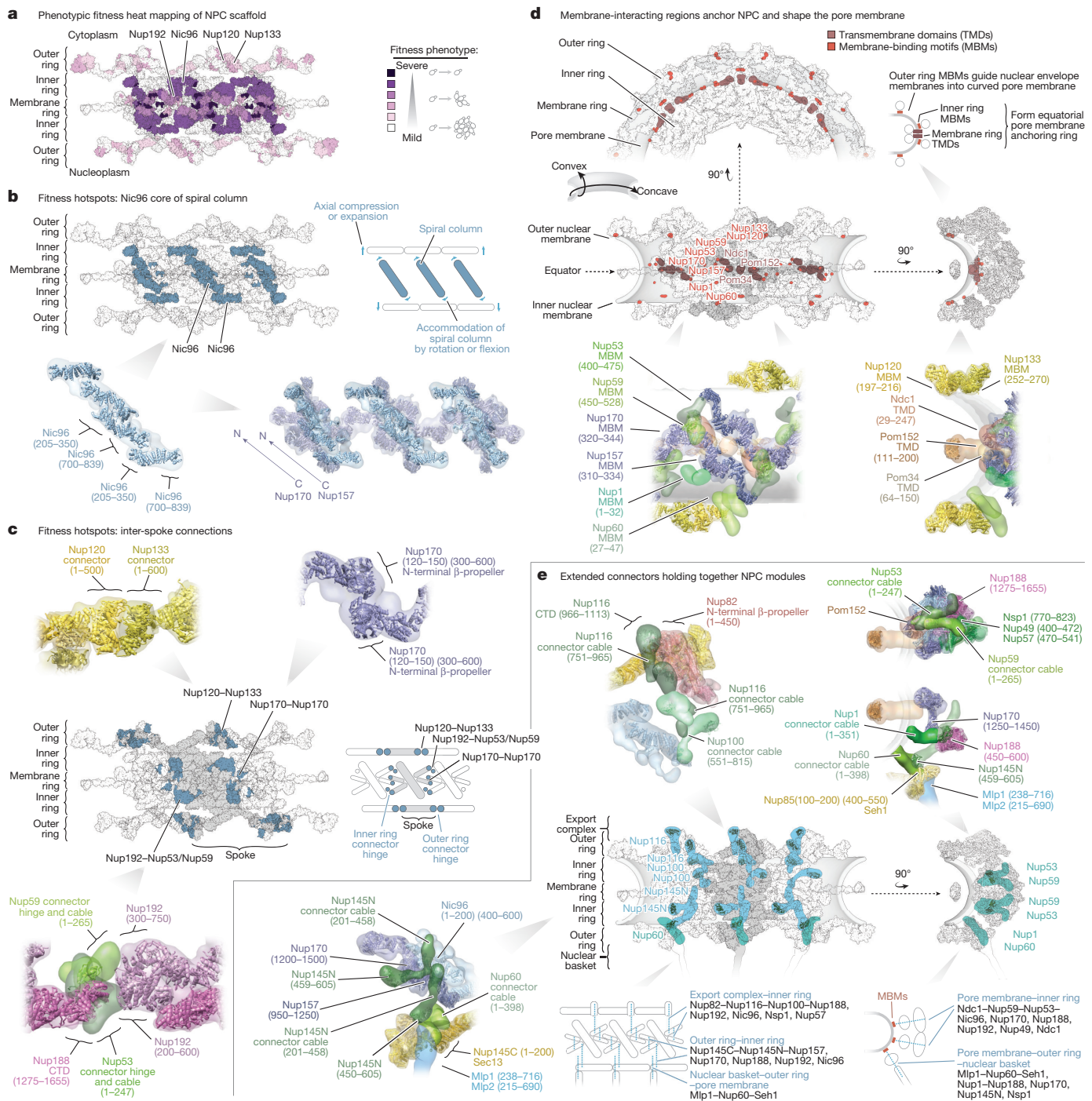


Figure 5 | Key NPC architectural features and principles. **a**, Severity of fitness defects, indicated in increasing intensity of purple shades for specific truncations of nucleoporins (Extended Data Fig. 9), mapped onto three spokes of the NPC. **b**, **c**, Structures corresponding to the position of the most severe defects (dark blue). In top panels in **b**, diagonally oriented columns reinforcing the core scaffold may accommodate NPC compression and expansion (diagram to right). Bottom, molecular details of Nup arrangement (relevant residue numbers indicated). In middle panels in **c**, the position of hotspots coincides with spoke-to-spoke connections. Central spoke, grey; flanking spokes, white (schematic, middle right). Top and bottom, molecular details of spoke-to-spoke connector hinges. **d**, Top left and centre left, three spokes shown as top and front views; centre right, one spoke in side view. Schematic indicates

convex and concave pore membrane curvatures. Positions of TMDs and MBMs are depicted and their proteins are labelled in brown and orange, respectively. Top right, diagrammatic side view showing how the MBMs and TMDs curve the pore membrane. Bottom, molecular details of the Nups containing the TMDs and MBMs. **e**, Second row left, three spokes in front view, showing how vertical connector Nups (cyan) spanning from the cytoplasmic to nuclear sides of the NPC connect the rings. Second row right, one spoke in side view, showing how horizontal connector Nups (aquamarine) connect modules spanning from the pore membrane to the central channel. First row and bottom row left show molecular details of the connectors within the NPC. Bottom row centre and right, diagrammatic views of the connectors depicted as blue dotted lines; modules connected labelled in blue; major Nups being contacted by connectors listed in grey.

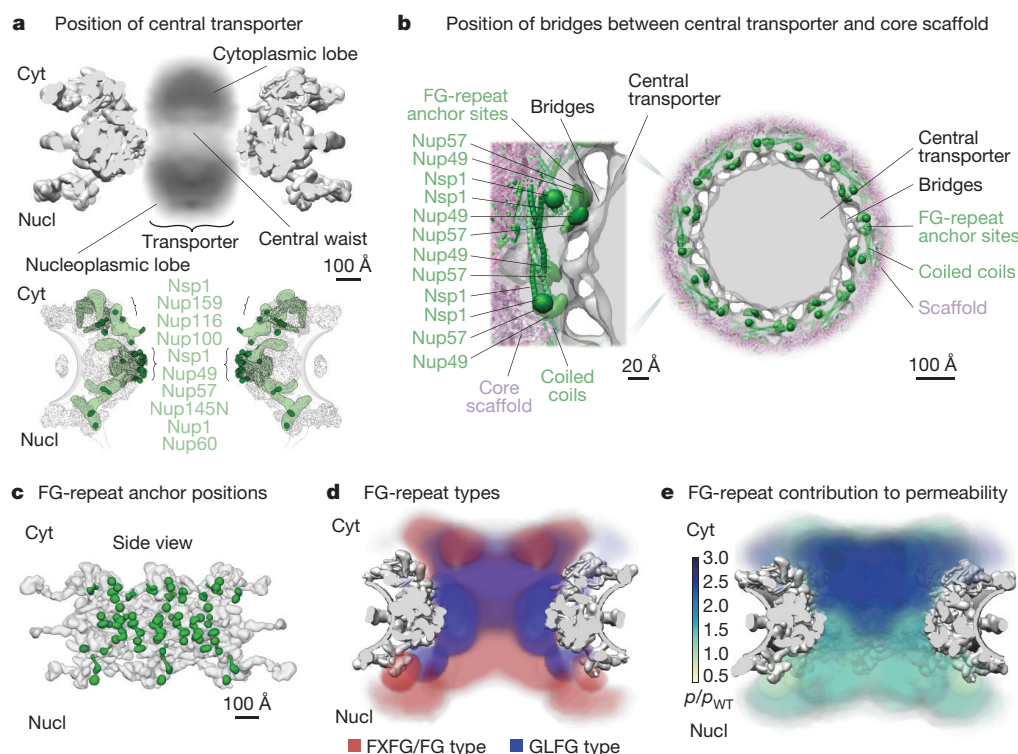


Figure 6 | The distribution of FG repeats informs the NPC transport gating mechanism. **a**, Central transporter density from the cryo-ET map (Fig. 3) is shown within the structure of the NPC scaffold (grey) (top). Features of the central transporter are indicated. Anchors (light green) in FG Nups largely direct the FG-repeat emanating points (dark green) towards the central channel (bottom). Scale bar, 100 Å. **b**, Central cross-section of the cryo-ET map (grey) with embedded representative NPC structure (Fig. 4), showing the central transporter and the bridges connecting it to the core scaffold in top view (scale bar, 100 Å), with a magnified view of one spoke on the left (scale bar, 20 Å). The anchor points for the FG repeats of Nup49, Nup57 and Nsp1 are depicted as

green densities. **c**, Position of FG-repeat anchor points (green) within a side view of three spokes of the scaffold (grey). Scale bar, 100 Å. **d**, Heat mapping of repeats of FXFG/FG type (red) and GLFG type (blue), from Brownian dynamics simulations (Methods), showing partitioning to different regions of the central channel. Scale bar, 100 Å. **e**, Heat mapping of the effect of FG-repeat region truncations on NPC permeability; the severity of the permeability defect³⁴ (measured as permeability relative to permeability in wild type, (p/p_{WT})) is indicated in increasing shades from minor (light green) to severe (dark blue). Cyt, cytoplasm; nucl, nucleoplasm. Scale bar, 100 Å.

reinforce the structural integrity of the NPC. Membrane-binding and transmembrane Nups are strategically placed throughout the core scaffold to stabilize pore membrane curvature and clamp the NPC to the nuclear envelope. Connectors run the length of each spoke, flexibly tying together all the major modules in the NPC. The architecture of the NPC is reminiscent of a suspension bridge, in which rigid supporting columns are firmly anchored to a substrate and flexible suspension cables connect the columns and roadway to provide a strong and resilient structure. We show that most FG Nup anchor points face inwards toward the NPC central channel to generate a highly concentrated milieu of FG repeats: FXFG/FG repeats form mRNA docking ‘traps’ at the entrance and exit of the channel, and GLFG repeats help form a cytoplasmically biased permeability barrier.

Despite differences, yeast and human NPCs retain a notable degree of structural conservation (Extended Data Fig. 12a, b). As a result, many of the conclusions drawn here should be applicable to the human NPC. To illustrate this point, we mapped the positions of yeast homologues of the oncogenic hotspot human Nup214, Nup98 and Tpr (ref. 2) (Extended Data Fig. 12c). Rather than being randomly scattered, these positions coincide with RNA-binding platforms on the cytoplasmic and nucleoplasmic faces of the NPC as well as with several critical connectors and associated FG regions. This conservation suggests that alterations in RNA export, and changes in NPC architecture induced by defective connectors, may underlie the altered behaviour of NPCs in cancer cells. Thus, our yeast structure provides a roadmap with the potential to advance our understanding of NPC physiology and nuclear transport in general.

Online Content Methods, along with any additional Extended Data display items and Source Data, are available in the online version of the paper; references unique to these sections appear only in the online paper.

Received 21 June 2017; accepted 6 February 2018.

Published online 14 March 2018.

1. Ptak, C., Aitchison, J. D. & Wozniak, R. W. The multifunctional nuclear pore complex: a platform for controlling gene expression. *Curr. Opin. Cell Biol.* **28**, 46–53 (2014).
2. Nofrini, V., Di Giacomo, D. & Mecucci, C. Nucleoporin genes in human diseases. *Eur. J. Hum. Genet.* **24**, 1388–1395 (2016).
3. Alber, F. *et al.* The molecular architecture of the nuclear pore complex. *Nature* **450**, 695–701 (2007).
4. Stanley, G. J., Fassati, A. & Hoogenboom, B. W. Biomechanics of the transport barrier in the nuclear pore complex. *Semin. Cell Dev. Biol.* **68**, 42–51 (2017).
5. Akey, C. W. & Goldfarb, D. S. Protein import through the nuclear pore complex is a multistep process. *J. Cell Biol.* **109**, 971–982 (1989).
6. Kosinski, J. *et al.* Molecular architecture of the inner ring scaffold of the human nuclear pore complex. *Science* **352**, 363–365 (2016).
7. Lin, D. H. *et al.* Architecture of the symmetric core of the nuclear pore. *Science* **352**, aaf1015 (2016).
8. Alber, F. *et al.* Determining the architectures of macromolecular assemblies. *Nature* **450**, 683–694 (2007).
9. Fernandez-Martinez, J. *et al.* Structure and function of the nuclear pore complex cytoplasmic mRNA export platform. *Cell* **167**, 1215–1228.e25 (2016).
10. Shi, Y. *et al.* Structural characterization by cross-linking reveals the detailed architecture of a coatomer-related heptameric module from the nuclear pore complex. *Mol. Cell. Proteomics* **13**, 2927–2943 (2014).
11. Briggs, J. A. Structural biology *in situ*—the potential of subtomogram averaging. *Curr. Opin. Struct. Biol.* **23**, 261–267 (2013).

12. Kim, S. J. *et al.* Integrative structure–function mapping of the nucleoporin Nup133 suggests a conserved mechanism for membrane anchoring of the nuclear pore complex. *Mol. Cell. Proteomics* **13**, 2911–2926 (2014).
13. Fernandez-Martinez, J. *et al.* Structure–function mapping of a heptameric module in the nuclear pore complex. *J. Cell Biol.* **196**, 419–434 (2012).
14. Herricks, T. *et al.* One-cell doubling evaluation by living arrays of yeast, ODELAY! G3 **7**, 279–288 (2017).
15. Aitchison, J. D., Rout, M. P., Marelli, M., Blobel, G. & Wozniak, R. W. Two novel related yeast nucleoporins Nup170p and Nup157p: complementation with the vertebrate homologue Nup155p and functional interactions with the yeast nuclear pore–membrane protein Pom152p. *J. Cell Biol.* **131**, 1133–1148 (1995).
16. Fischer, J., Teimer, R., Amlacher, S., Kunze, R. & Hurt, E. Linker Nups connect the nuclear pore complex inner ring with the outer ring and transport channel. *Nat. Struct. Mol. Biol.* **22**, 774–781 (2015).
17. von Appen, A. *et al.* *In situ* structural analysis of the human nuclear pore complex. *Nature* **526**, 140–143 (2015).
18. Marelli, M., Lusk, C. P., Chan, H., Aitchison, J. D. & Wozniak, R. W. A link between the synthesis of nucleoporins and the biogenesis of the nuclear envelope. *J. Cell Biol.* **153**, 709–724 (2001).
19. Vollmer, B. *et al.* Dimerization and direct membrane interaction of Nup53 contribute to nuclear pore complex assembly. *EMBO J.* **31**, 4072–4084 (2012).
20. Seo, H. S. *et al.* Structural and functional analysis of Nup120 suggests ring formation of the Nup84 complex. *Proc. Natl Acad. Sci. USA* **106**, 14281–14286 (2009).
21. Drin, G. *et al.* A general amphipathic α -helical motif for sensing membrane curvature. *Nat. Struct. Mol. Biol.* **14**, 138–146 (2007).
22. Mészáros, N. *et al.* Nuclear pore basket proteins are tethered to the nuclear envelope and can regulate membrane curvature. *Dev. Cell* **33**, 285–298 (2015).
23. Upla, P. *et al.* Molecular architecture of the major membrane ring component of the nuclear pore complex. *Structure* **25**, 434–445 (2017).
24. Meinema, A. C. *et al.* Long unfolded linkers facilitate membrane protein import through the nuclear pore complex. *Science* **333**, 90–93 (2011).
25. Knockenhauer, K. E. & Schwartz, T. U. The nuclear pore complex as a flexible and dynamic gate. *Cell* **164**, 1162–1171 (2016).
26. Folkmann, A. W., Noble, K. N., Cole, C. N. & Wentz, S. R. Dbp5, Gle1–IP6 and Nup159: a working model for mRNP export. *Nucleus* **2**, 540–548 (2011).
27. Saroufim, M. A. *et al.* The nuclear basket mediates perinuclear mRNA scanning in budding yeast. *J. Cell Biol.* **211**, 1131–1140 (2015).
28. Meseroll, R. A. & Cohen-Fix, O. The malleable nature of the budding yeast nuclear envelope: flares, fusion, and fenestrations. *J. Cell. Physiol.* **231**, 2353–2360 (2016).
29. Eibauer, M. *et al.* Structure and gating of the nuclear pore complex. *Nat. Commun.* **6**, 7532 (2015).
30. Paradise, A., Levin, M. K., Korza, G. & Carson, J. H. Significant proportions of nuclear transport proteins with reduced intracellular mobilities resolved by fluorescence correlation spectroscopy. *J. Mol. Biol.* **365**, 50–65 (2007).
31. Adams, R. L. & Wentz, S. R. Uncovering nuclear pore complexity with innovation. *Cell* **152**, 1218–1221 (2013).
32. Patel, S. S., Belmont, B. J., Sante, J. M. & Rexach, M. F. Natively unfolded nucleoporins gate protein diffusion across the nuclear pore complex. *Cell* **129**, 83–96 (2007).
33. Adams, R. L., Terry, L. J. & Wentz, S. R. Nucleoporin FG domains facilitate mRNP remodeling at the cytoplasmic face of the nuclear pore complex. *Genetics* **197**, 1213–1224 (2014).
34. Timney, B. L. *et al.* Simple rules for passive diffusion through the nuclear pore complex. *J. Cell Biol.* **215**, 57–76 (2016).
35. Yamada, J. *et al.* A bimodal distribution of two distinct categories of intrinsically disordered structures with separate functions in FG nucleoporins. *Mol. Cell. Proteomics* **9**, 2205–2224 (2010).
36. Devos, D. *et al.* Components of coated vesicles and nuclear pore complexes share a common molecular architecture. *PLoS Biol.* **2**, e380 (2004).
37. Faini, M., Beck, R., Wieland, F. T. & Briggs, J. A. Vesicle coats: structure, function, and general principles of assembly. *Trends Cell Biol.* **23**, 279–288 (2013).
38. Rout, M. P. & Field, M. C. The evolution of organellar coat complexes and organization of the eukaryotic cell. *Annu. Rev. Biochem.* **86**, 637–657 (2017).
39. Obado, S. O. *et al.* Interactome mapping reveals the evolutionary history of the nuclear pore complex. *PLoS Biol.* **14**, e1002365 (2016).
40. Iwamoto, M. *et al.* Compositionally distinct nuclear pore complexes of functionally distinct dimorphic nuclei in the ciliate *Tetrahymena*. *J. Cell Sci.* **130**, 1822–1834 (2017).
41. Obado, S. O., Field, M. C. & Rout, M. P. Comparative interactomics provides evidence for functional specialization of the nuclear pore complex. *Nucleus* **8**, 340–352 (2017).
42. Bui, K. H. *et al.* Integrated structural analysis of the human nuclear pore complex scaffold. *Cell* **155**, 1233–1243 (2013).
43. Deblor, E. W. *et al.* A fence-like coat for the nuclear pore membrane. *Mol. Cell* **32**, 815–826 (2008).

Supplementary Information is available in the online version of the paper.

Acknowledgements We thank B. Webb (UCSF) for help with the Integrative Modelling Platform, the Rockefeller University Outreach Program for support for A.S.C., the NYULMC OCS Microscopy Core, K. Uryu and the EMRC Resource Center (Rockefeller University) for assistance with negative-stain electron microscopy, F. Alber, M. C. Field, N. Ketaren, S. Obado, R. Hayama and D. Simon for feedback and critical reading of the manuscript, and L. Herlands for support and encouragement. The work was supported by a NSF GRF 1650113 (I.E.C.), a NSF grant CHE-1531823 (M.F.J.), the SIMR (J.L.G.), NIH grants R01 GM080477 (J.L.G.), U54 GM103511 (B.T.C., A.S., J.D.A. and M.P.R.), R01 GM112108 (M.P.R. and J.D.A.), P41 GM109824 (M.P.R., A.S., J.D.A. and B.T.C.), P50 GM076547 (J.D.A.), R01 GM063834 (C.W.A.), R01 GM080139 (S.J.L.), P41 GM103314 (B.T.C.), R01 GM083960 (A.S.) and U54 DK107981 (M.P.R. and J.D.A.). We are grateful for the support provided by G. Blobel, who inspired the work presented here.

Author Contributions The order of first co-authors was determined through a random selection process. I.N., J.F.-M., A.S.C., R.W. and M.P.R. performed the affinity purifications; W.Z., J.F.-M., R.W., R.M., E.Y.J., M.P.R. and B.T.C. performed the quantitative mass spectrometry; M.S., B.D.S., J.R.U. and J.L.G. performed the calibrated imaging; J.A.H., B.T.C. and M.F.J. performed the charge detection mass spectrometry; Y.S., J.F.-M., R.W., I.N., J.W. and B.T.C. performed the chemical crosslinking with mass spectrometry; C.W.A., S.J.L., I.N., Z.Y. and M.J.d.I.C. performed the cryo-ET; S.J.K. performed the small-angle X-ray scattering; T.H., J.F.-M. and J.D.A. performed the phenotypic profiling; P.U. and D.L.S. performed the negative-stain electron microscopy; S.J.K., B.R., I.E.C., R.P., I.E., C.H.G. and A.S. performed the integrative structure computations; S.J.L., C.W.A., B.T.C., A.S. and M.P.R. supervised the project; S.J.K., J.F.-M., I.N., Y.S., W.Z., B.R., S.J.L., C.W.A., B.T.C., A.S. and M.P.R. wrote the manuscript.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Publisher's note: Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations. Correspondence and requests for materials should be addressed to M.P.R. (rout@rockefeller.edu), B.T.C. (chait@rockefeller.edu), A.S. (sali@salilab.org), C.W.A. (cakey@bu.edu) or S.J.L. (sludtke@bcm.edu).

METHODS

1. Yeast strains and materials. All *S. cerevisiae* strains used in this study are listed in Supplementary Table 5, with the exception of the Nup84 complex truncation mutants¹³ and the Pom152 truncation mutants²³. Unless otherwise stated, strains were grown at 30 °C in YPD medium (1% yeast extract, 2% bacto-peptone and 2% glucose). The diploid *S. uvarum* strain (ATCC 9080) was grown and processed for nuclear envelope purification as previously described⁴⁴.

The following materials were used in this study: Dynabeads M-270 Epoxy (143.02D; Invitrogen); rabbit IgG (55944; MP Biomedicals); protease inhibitor cocktail (P-8340; Sigma-Aldrich); and Solution P (2 mg Pepstatin A, 90 mg PMSE, and 5 ml of absolute ethanol).

2. Immuno-purification of the endogenous *S. cerevisiae* NPC. An immuno-purification protocol for the isolation of endogenous whole NPCs from *S. cerevisiae* was developed using previously published methodology^{3,45–49}. *S. cerevisiae* Mlp1-, Nup84- or Nup82-encoding genes were genomically tagged with PrA preceded by the human rhinovirus 3C protease (PPX) target sequence (GLEVLFGPS). Cells were grown in YPD medium at 30 °C until early log phase ($\sim 2 \times 10^7$ cells/ml), collected, frozen in liquid nitrogen and cryogenically lysed in a planetary ball mill PM 100 (Retsch) (<http://lab.rockefeller.edu/rout/protocols>). Frozen cell powder was resuspended in 9 volumes of resuspension buffer (20 mM HEPES–KOH pH 7.4, 50 mM potassium acetate, 20 mM NaCl, 2 mM MgCl₂, 0.5% (w/v) Triton X-100, 0.1% (w/v) Tween-20, 1 mM DTT, 10% (v/v) glycerol, 1/500 (v/v) protease inhibitor cocktail (Sigma)). Cell lysate was clarified by centrifugation at 2,500 RCF for 5 min followed by filtration through 1.6- μ m filters (Whatman glass microfibre syringe filters). Magnetic beads (Invitrogen) conjugated to rabbit IgG antibodies (<http://lab.rockefeller.edu/rout/protocols>) were added to the clarified cell lysate at a concentration of 50 μ l slurry per 1 g of frozen cell powder and incubated for 30 min at 4 °C. Beads were washed once with 1 ml of elution buffer without protease inhibitors (20 mM HEPES–KOH pH 7.4, 50 mM potassium acetate, 20 mM NaCl, 2 mM MgCl₂, 0.1% (w/v) Tween-20, 1 mM DTT, 10% (v/v) glycerol). For native elution of the complex, the desired volume of elution buffer with PreScission protease (GE Healthcare) (1/15 (v/v)) was added to the beads and incubated for 45 min at 4 °C. A magnet was used to remove the beads and collect the supernatant. Beads were subsequently washed with the desired volume of elution buffer containing 1/500 (v/v) protease inhibitor cocktail (Sigma). The total elution volume was centrifuged at 20,000g for 5 min to remove the residual magnetic beads. Typical yield of the immuno-purification is $\sim 4 \mu$ g of isolated NPCs per 1 g frozen cell powder (see Extended Data Fig. 2b for SDS-PAGE analysis; for gel source data, see Supplementary Fig. 1).

3. Mass and stoichiometry of the native *S. cerevisiae* NPC. Quantification of the absolute stoichiometry of each nucleoporin in the native NPCs was performed using a strategy that combined several orthogonal methods (Extended Data Fig. 2a): (1) use of synthetic concatemers of tryptic peptides or QconCATs⁵⁰ to define the relative stoichiometry of each component by quantitative mass spectrometry in affinity-captured NPCs; (2) *in vivo* calibrated imaging analysis of GFP-tagged Nups⁵¹, to quantify the absolute copy number per NPC of Nups selected to represent each major module of the NPC; and (3) charge detection mass spectrometry to measure the total mass of affinity-captured NPCs⁵². For the calculation of the integrative NPC structure, the final copy numbers were rounded to fit the known NPC C₈-symmetry and these values are indicated in Supplementary Table 2a.

3.1 NPC QconCAT design and purification. Mass spectrometry quantification of the relative amounts of each nucleoprotein in the purified NPC complex was performed using two specifically designed, heavy-labelled synthetic internal standards or QconCATs^{50,53} (Extended Data Fig. 2d, e) formed by concatenated quantotypic nucleoporin peptides. To minimize the potential effect of having different residues flanking the trypsin cleavage site on the cleavage efficiency, we included the native three-residue flanking sequences framing the trypsin cleavage site for each peptide⁵⁴. For QconCAT-A (Extended Data Fig. 2d), two peptides for each of the nucleoporins and one peptide for *Staphylococcus aureus* protein A and *Aequorea victoria* GFP proteins were selected (Supplementary Table 7) on the basis of their favourable signal responses in liquid chromatography–mass spectrometry analyses of NPC samples and by fitting to the following criteria (when possible): (1) the native three-residue flanking sequences at both sides of the trypsin cleavage sequence do not contain Lys or Arg; (2) avoid the presence of Cys or Met residues within the peptide; (3) avoid the presence of potential internal trypsin cleavage sites (Lys or Arg residues); (4) peptides should be less than 3,000 Da (small size); and (5) avoid peptides showing obvious interferences from co-eluting peptides during the liquid chromatography separation for mass spectrometry analyses. QconCAT-B included two quantotypic peptides for Nup159, Mlp2, Nup192, Nup84, Nup85, Nup120, Nup49, Nup57, Pom152 and Nic96, and the same GFP peptide as in QconCAT-A (Supplementary Table 7). As an internal control, both QconCAT-A

and -B included the same peptides for Nic96, Pom152 and GFP. Each synthetic gene was designed by concatenation of the sequences encoding the selected peptides and addition of a 6 \times His C-terminal tag (Extended Data Fig. 2d). A 3 \times FLAG peptide was also included at the N terminus of QconCAT-A, resulting in a protein of 148.2 kDa. The *Escherichia coli* codon optimized sequences were cloned into: (1) plasmid pET15-b (as a NcoI–XhoI fragment) in the case of QconCAT-A; and (2) pGEX6p-1 (as a BamHI–XhoI fragment) in the case of QconCAT-B, resulting in the expression of a 68.1 kDa protein with a N-terminal GST tag that was mainly used as a sacrificial peptide⁵⁵. The QconCAT proteins were expressed by growing 300 ml of BL21 *E. coli* cells at 37 °C to OD₆₀₀ = 0.6 in minimal M9 medium without ammonium chloride^{50,53} supplemented with light amino acids and 0.5 mg/ml of heavy arginine and lysine (L-arginine:HCl ¹³C₆; L-lysine:2HCl ¹³C₆, Cambridge Isotope Laboratories). IPTG (1 mM) was used to induce expression of the constructs for 3 h at 37 °C. Collected cells were processed using BugBuster Extraction Reagent (Novagen), as indicated by the manufacturer, to isolate the inclusion bodies where the QconCAT protein is accumulated. The full-length QconCAT-A was then purified by resuspending the inclusion bodies pellet in binding buffer (20 mM sodium phosphate pH 7.4, 45 mM imidazole, 500 mM NaCl, 6 M guanidinium chloride, 10 mM TCEP (0.5 M Bond-Breaker TCEP solution, Thermo Fisher Scientific), 1/500 protease inhibitor cocktail (Roche)) and passed through an equilibrated His-Trap HP (GE Healthcare) at room temperature. The retained NPC QconCAT-A was then eluted in 20 mM sodium phosphate pH 7.4, 500 mM imidazole, 500 mM NaCl, 6 M guanidinium hydrochloride, 1 mM TCEP, 1/500 protease inhibitor cocktail. One-hundred-microlitre aliquots of the resulting elution were precipitated to eliminate the guanidinium hydrochloride by adding ice-cold ethanol to a final concentration of 90% and incubating the samples at –20 °C for 2 h. Samples were then centrifuged for 10 min at 14,000 r.p.m. and 4 °C to pellet the precipitated protein. The resulting pellet was washed with ice-cold 90% ethanol and allowed to air-dry until most of the liquid was evaporated, leaving a wet pellet. These pellets were solubilized with 5% SDS, 500 mM Tris–HCl pH 8.0, 5 mM TCEP buffer, by incubating for 5 min at room temperature and 5 min at 72 °C and centrifuged for 10 min at 14,000 r.p.m. at room temperature. The supernatants were recovered and two of them combined and injected into a TSKgel G4000SWxl size-exclusion column (TOSOH Bioscience) coupled to a TSKgel SWxl guard column (TOSOH Bioscience), pre-equilibrated in running buffer (40 mM HEPES–KOH pH 7.0, 150 mM NaCl, 0.1% SDS, 5 mM TCEP, 1 mM EDTA). Two-hundred-microlitre fractions were collected and analysed by SDS-PAGE to detect the presence of the QconCAT-A peak. Fractions containing the full-length pure protein were supplemented with a final 20% glycerol (v/v), aliquoted, flash-frozen in liquid nitrogen and stored at –80 °C for further use. In the case of QconCAT-B, the protein was purified using His-Trap HP and the elution precipitated and prepared as described for QconCAT-A. The resulting sample was injected into a TSKgel Super SW3000 size-exclusion column (TOSOH Bioscience) pre-equilibrated in running buffer (40 mM HEPES–KOH pH 7.0, 150 mM NaCl, 0.1% SDS, 5 mM TCEP, 1 mM EDTA). One-hundred-microlitre fractions were collected and analysed by SDS-PAGE to detect the presence of the QconCAT-B peak. Fractions containing the full-length pure protein were stored as indicated for QconCAT-A.

For the quantitative mass spectrometry analysis, the native NPCs from *S. cerevisiae* PPX–PrA-tagged haploid strains were affinity captured as described above, or purified as enriched NPCs from a diploid *S. uvarum* strain using a subfractionation method previously described in detail^{44,56–58} (<http://lab.rockefeller.edu/rout/protocols>), using 0.035 mg heparin per mg of fraction protein. For affinity-captured NPCs, the natively eluted NPCs (5 μ g) were concentrated by pelleting at 40,000 r.p.m. for 20 min at 4 °C in a TLA 55 rotor (Beckman). In the case of subfractionation-enriched NPCs, a volume of the 1.45 M:1.85 M sucrose gradient fraction that contained an estimated 5 μ g of NPCs was diluted 1/5 (v/v) in bt-DMSO buffer (10 mM bis-Tris-HCl pH 6.5, 0.1 mM MgCl₂, 20% DMSO) and pelleted at 15,000 r.p.m. for 450 min at 4 °C in a TLA 55 rotor (Beckman). For in-solution mass spectrometry analysis of subfractionation-enriched NPCs, 0.1 μ g of QconCAT-A were immobilized on Dynabeads His-Tag Isolation and Pull-down resin (Thermo Fisher Scientific) pre-equilibrated in binding buffer (20 mM HEPES, 150 mM NaCl, 8 M urea, 5 mM TCEP). The purified protein sample was incubated with the resin for 20 min at room temperature, and washed with binding buffer 5 \times 200 μ l to eliminate residual SDS; in-solution and in-gel analyses showed consistent results (not shown), so most of the further analyses were performed in-gel to improve consistency, speed and throughput. For the solid-state in-gel mass spectrometry analyses, pelleted NPCs were solubilized in 10 μ l of 0.5 M Tris–HCl pH 8.0, 5% SDS by incubating at 72 °C for 5 min and then diluted 1:1 with 20% glycerol, 50 mM TCEP, 0.5 mM EDTA, 0.05% (w/v) bromophenol blue. Approximately equimolar amounts of 0.1 μ g of purified QconCAT-A or 0.045 μ g of purified QconCAT-B were added to each 5- μ g NPC sample. Samples were then incubated at 72 °C for 10 min,

cooled to room temperature and treated with a final 30 mM of iodoacetamide (Sigma), at room temperature in the dark for 30 min. Samples were then loaded into a 4% (37.5:1) stacking acrylamide SDS-PAGE gel prepared in-house. The resulting bands, containing a mixture of whole NPCs and QconCAT protein (labelled with a stable isotope), were excised and processed for quantitative mass spectrometry analyses.

3.2 Mass spectrometry characterization of QconCAT labelled with a stable isotope. The mass of purified intact QconCAT-A protein, labelled with a stable isotope, was analysed by matrix-assisted laser desorption/ionization (MALDI) (Extended Data Fig. 2e) on a JEOL JMS-S3000 SpiralTOF mass spectrometer using the ultra-thin-layer sample preparation method^{59,60} in which α -cyano-4-hydroxycinnamic acid (Sigma) was used as the matrix. The mass of QconCAT-A was internally calibrated with horse myoglobin. Mass calibration and background subtraction were carried out with the JEOL msTornado control software, and additional analyses were carried out with the MoverZ software⁶¹. The QconCAT-A protein was also characterized by peptide mapping, in which tryptic peptides from in-gel digestion were loaded onto a PicoFrit column (New Objective) with an integrated emitter tip (360-mm O.D., 50-mm I.D., 10-mm tip) self-packed with 6 cm of reverse-phase C18 material (ReproSil-Pur C18-AQ, 3-mm beads, Dr. Maisch GmbH), and analysed with a LTQ Orbitrap Velos mass spectrometer (Thermo Fisher Scientific), with a Agilent 1200 series HPLC system (Agilent) and a micro electrospray source built in-house. The purified QconCAT-B was characterized by peptide mapping on a Thermo Orbitrap Fusion mass spectrometer, with a Thermo Easy-nLC 1000 HPLC and a Thermo Easy-Spray electrospray source.

3.3 Stoichiometry quantification of NPC using QconCAT and by mass spectrometry. Mixtures of yeast NPC proteins and QconCAT, labelled with a stable isotope, were enzymatically digested either in solution in the presence of urea or inside a SDS-PAGE gel matrix. For in-solution digestion, a mixture of the NPCs and immobilized QconCATs on His-Dynabeads were sequentially digested at room temperature by Endoproteinase Lys-C in 8 M urea for 66 h and by trypsin in 2 M urea for 3 h. For in-gel digestion, proteins in the gel matrix were digested in 100 mM Tris-HCl at room temperature either sequentially by 0.25–2 μ g Endoproteinase Lys-C for 66 h and by 3–25 μ g trypsin for 3 h, or—in later experiments—by 25 μ g trypsin alone for 3 h. The resulting peptides were analysed in duplicate by liquid chromatography–mass spectrometry using a Thermo Fusion or a Thermo Q Exactive Plus mass spectrometer, with a Thermo Easy-nLC 1000 HPLC and a Thermo Easy-Spray electrospray source. The ratios of light nucleoporin (L) to heavy QconCAT proteins (H) for standard peptides were obtained using MaxQuant⁶², complemented with manual determination.

We incorporated two standard peptides from each nucleoporin into the QconCAT standard to enable us to check for internal consistency of the measured L/H ratios for each nucleoporin. Our check required that the relative standard deviations of L/H ratios for two standard peptides from two duplicate liquid chromatography–mass spectrometry runs—that is, for a total of four measurements per nucleoporin—be $\leq 25\%$. When deriving relative stoichiometry for any given preparation of NPCs analysed in different replication experiments, we corrected for variations in the mixing ratio of light nucleoporins and heavy QconCAT proteins by scaling the measured L/H ratios to minimize the sum of the relative standard deviations of the L/H ratios over all nucleoporins. The resulting scaled L/H ratios from different experiments were used to derive the average L/H ratios and standard deviations. To assay for potential nucleoporin stoichiometry bias arising from capture through particular affinity handles, we used stable isotope labelling with amino acids in cell culture followed by mass spectrometry (SILAC–MS) analysis of these preparations versus the nuclear envelope preparation. We performed $n = 2$ or 3 technical and 2 biological replicas for NPCs purified by subfractionation procedures from a diploid, non-tagged *S. uvarum* strain, and $n = 1$ –3 technical and 4 biological replicas for the nuclear-envelope-corrected affinity-captured NPCs from haploid, tagged *S. cerevisiae* strains (Fig. 1a).

The absolute stoichiometry (Fig. 1 and Supplementary Table 2a) was then determined by normalizing the summed copies of Nup188, Nup120 and Nic96 per NPC to 64 copies (that is, 16 for Nup188 and Nup120, and 32 for Nic96).

3.4 SILAC-MS analyses of the NPC stoichiometry. A preparation of yeast nuclear envelopes obtained by a previously established subfractionation method⁴⁴ does not involve disruption of the nuclear envelope membrane by detergents and generates sheets of nuclear envelope studded with intact NPCs. To assess the degree to which the affinity-captured NPCs were intact, we used SILAC-MS to compare the levels of each Nup in the affinity-captured preparation relative to those in the nuclear envelope preparation. To do this, the nuclear envelope sample labelled with light isotopes was mixed with Mlp1-PPX-PrA affinity-captured NPC sample labelled with a heavy isotope (L-lysine:2HCL $^{13}\text{C}_6$) in a SILAC experiment. Mixtures of nuclear envelope proteins and NPCs labelled with stable isotopes, purified using the Mlp1-PPX-PrA handle, were digested sequentially in gel matrix

by Endoproteinase LysC and by trypsin. Resulting peptides were analysed by liquid chromatography–mass spectrometry on a Thermo Q Exactive Plus mass spectrometer, with a Thermo Easy-nLC 1000 HPLC and a Thermo Easy-Spray electrospray source. H/L ratios for all peptides were obtained using MaxQuant⁶², complemented with manual examinations. The H/L ratios of peptides were used to derive the H/L ratios of nucleoporins and standard deviations (data not shown). The result showed that the affinity-capture process does not affect the overall ratios of the major Nups and NPC modules relative to the nuclear envelope samples (data not shown), indicating that the affinity-capture procedure generates intact NPCs. We also used this comparison relative to nuclear envelopes to correct for the slight increases observed in the ratios of Nups closely associated with the Mlp1 handle in the affinity-captured NPCs (Fig. 1a).

3.5 In vivo calibrated imaging analysis of GFP-tagged Nups. Calibrated imaging data were acquired as previously described⁵¹. Using the avalanche photodiode imaging module of a Zeiss confocal 3, confocal z-stacks of live yeast were acquired with a 40×1.2 NA Plan-Apochromat water objective. The 488-nm laser line was used to excite GFP, with a 405/488/561 dichroic. Emission was reflected with a LP580 emission dichroic and collected through a BP 505–540-nm emission filter. The pinhole was set to 1 Airy unit. The zoom was set to maintain a pixel size of 55 nm, and a z-step size of 400 nm was used. After acquisition, images were binned in XY by 2, resulting in an effective pixel size of 110 nm, and anaphase cells were analysed for diffraction-limited Nup spots along the anaphase bridge. These spots, when present, were fit to a 2D Gaussian curve to obtain the amplitude of the signal. The z-slice with the maximum signal intensity of the spot was analysed. Fluorescence correlation spectroscopy was used to convert the amplitude of the Gaussian fit of the Nup spot number of molecules of GFP. In brief, using a strain expressing only cytosolic GFP, fluorescence correlation spectroscopy determined the average number of molecules in the focal volume, as previously described⁵¹. Then, the amplitude of the signal of the Nup spot was compared to the intensity of cytosolic GFP, taken with the same imaging setup. For all measurements, number 1.5 coverslips were measured for uniformity, and the correction collar of the water objective was optimized for this thickness using signal intensity of Alexa Fluor 488 in solution. For each day data were acquired, the calibration using cytosolic monomer GFP was obtained.

3.6 Phospholipid analysis. These analyses were performed by Avanti Polar Lipids using their standard protocols.

3.7 Label-free mass spectrometry quantification of the NPC and associated proteins. Raw mass spectrometry files from QconCAT Mlp1-PPX-PrA immuno-isolation experiments were analysed using the MaxQuant iBAQ method⁶³. Only peptides that were not isotopically labelled (that is, not QconCAT) were considered. Proteins were filtered to require more than three unique peptides per protein, and stoichiometries normalized to the absolute minimum value of the difference between label-free and the QconCAT stoichiometry for all the Nups (Extended Data Fig. 3c and Supplementary Table 8). Stoichiometries were multiplied by molecular weight to obtain mass per NPC complex and the results summed to obtain total mass of the NPC (Fig. 1c and Extended Data Fig. 3c).

3.8 Living mass of the NPC with charge detection mass spectrometry. The charge detection mass spectrometry instrument has previously been described^{52,64}. In brief, the measurements are made by trapping single ions in a linear electrostatic ion trap. As the ions oscillate back and forth in the trap, they pass through a cylindrical electrode. The charge induced on the electrode is detected by a charge sensitive preamplifier. The resulting signal is amplified and digitized, and then analysed using fast Fourier transforms. The fundamental frequency provides the m/z and the magnitude is proportional to the charge. The mass of each ion is then obtained by multiplying the charge and m/z . Each NPC sample was characterized by measuring the masses of several thousand ions individually and then binning the masses to yield a true mass spectrum (Fig. 1b, c).

4. Chemical cross-linking and mass spectrometry analysis of the cross-linked NPC. NPCs were immuno-purified from Mlp1-PPX-PrA, Nup82-PPX-PrA and Nup84-PPX-PrA *S. cerevisiae* strains. After native elution, 0.5 or 1.0 mM disuccinimidyl suberate (DSS) was added and sample incubated at room temperature for 30 min with gentle shaking ($\sim 1,000$ r.p.m.). The reaction was quenched with 50 mM ammonium bicarbonate or SDS-PAGE buffer containing 100 mM Tris-HCl. The sample was then precipitated using 90% methanol at -80°C or concentrated in a speed vacuum before separation by SDS electrophoresis.

The sample was reduced by 10 mM tris-(2-carboxyethyl)-phosphine (Invitrogen) at 80°C for 15–20 min, cooled to room temperature and alkylated by 50 mM iodoacetamide for 20 min in the dark to block the formation of disulfide bonds. After reduction and alkylation, the cross-linked complexes were separated by 3–8% SDS-PAGE (NuPAGE Tris-Acetate Fisher) to reduce the complexity of the sample. For in-gel digestion, the high-molecular-weight-region gel bands (>460 kDa, estimated by the high-molecular-weight protein markers, Invitrogen)

corresponding to the cross-linked NPC proteins were sliced and proteolysed by trypsin as previously described^{10,65}. In brief, gel plugs were crushed into small pieces, ~5–10 µg of sequencing-grade trypsin (Promega) per ~100 µg protein was added with subsequent 6–8 h incubation. This proteolysis was repeated once more to ensure optimal results. Peptides were extracted by formic acid and acetonitrile, desalted on C18 cartridges (Waters) and snap-frozen before fractionation.

To reduce the complexity of the sample, proteolysed mixtures were separated by an orthogonal two-step fractionation strategy. First, size exclusion chromatography⁶⁶ was used for size-based separation of peptides into 2–4 fractions (~2–10 kDa). Then, a secondary fractionation using a self-packed basic (at pH 10) C18 resins (Dr. Maisch GmbH) resulted in 10–12 peptide fractions, which were subsequently analysed by liquid chromatography–mass spectrometry.

Each peptide fraction was dissolved in the sample loading buffer (5% MeOH and 0.2% FA) and analysed either by an Orbitrap Q Exactive Plus mass spectrometer or a LTQ Velos Orbitrap Pro mass spectrometer (Thermo Fisher Scientific). The Q Exactive Plus instrument was directly coupled to an easy-nLC system (Thermo Fisher Scientific) for electrospray. The cross-linked peptides were loaded onto the Easy-Spray columns (15-cm prepacked columns that are filled with C18 reverse-phase material of 2 or 3 µm particle size, 200 Å pore size and 50 µm inner diameter, Thermo Fisher Scientific) that were heated to 35 °C. Mobile phase A consisted of 0.1% formic acid and mobile phase B of 100% ACN with 0.1% formic acid. Peptides were eluted in liquid chromatography gradients of 120 min (for example, a liquid chromatography gradient of 3–7% B, 0–6 min; 7–28% B, 6–101 min; 28–100% B, 101–113 min; followed by equilibration with 100% A until 120 min). Flow rates were set at ~250–275 nL/min. Other instrumental parameters for chemical cross-linking and mass spectrometry analyses include: capillary temperature: 250–275 °C; target mass resolutions (at 200 Th): 70,000 for mass spectrometry and 17,500 for tandem mass spectrometry; AGC targets: $1-3 \times 10^6$ (full mass) and 2×10^5 (tandem mass spectrometry); mass spectrometry range of 300–1,700 Th; isolation window: 1.3–1.7 Th; higher-energy collisional dissociation normalized energy: 24–29; dynamic exclusion allowed once per 75–90 s. The top 8 most abundant ions (with charge stage of 3–7 and intensity thresholds of 3,000–7,500 ions) were selected for fragmentation by higher-energy collisional dissociation. The maximum injection times were set at 200 ms (for mass spectrometry) and 500–800 ms (for tandem mass spectrometry). For samples that were analysed by Orbitrap Velos, the cross-linked peptide mixtures were pressure-loaded onto a self-packed PicoFrit column with integrated electrospray ionization emitter tip (360 O.D., 75 I.D. with 15 µm tip, New Objective). The column was packed with 10–15 cm reverse-phase C18 material (3 µm porous silica, 200 Å pore size, Dr. Maisch GmbH). Mobile phase A consisted of 0.5% acetic acid and mobile phase B of 70% ACN with 0.5% acetic acid. The peptides were eluted in a 120- or a 140-min liquid chromatography gradient (8% B to 50% B, 0–93 min, followed by 50% B to 100% B, 93–110 min and equilibrated with 100% A until 120 or 150 min) using a HPLC system (Agilent), and analysed with a LTQ Velos Orbitrap Pro mass spectrometer using similar parameters to the Q Exactive Plus instrument.

The raw data were searched by pLink⁶⁷ using a FASTA database containing 34 NPC protein sequences. An initial MS1 search window of 5 Da was allowed to cover all isotopic peaks of the cross-linked peptides. The data were automatically filtered using a mass accuracy of $MS1 \leq 10$ p.p.m. (parts per million) and $MS2 \leq 20$ p.p.m. of the theoretical monoisotopic (A_0) and other isotopic masses ($A + 1$, $A + 2$, $A + 3$ and $A + 4$) as specified in the software. Other search parameters included cysteine carbamidomethyl as a fixed modification and methionine oxidation as a variable modification. A maximum of two trypsin missed-cleavage sites was allowed. The initial search results were obtained using a default 5% false discovery rate expected by target–decoy search strategy. All spectra were manually verified as previously described^{9,10,65,68,69}. The cross-linking data were analysed and plotted by an online software tool, CX-Circos (<http://cx-circos.net>; W.J. *et al.*, manuscript in preparation) (Fig. 2).

5. Cryo-electron tomography of whole NPCs. We used cryo-ET and sub-tomogram averaging (Extended Data Fig. 4a) to obtain a final map with a global resolution of 28 Å; the inner ring was solved at 20–25 Å (Extended Data Fig. 5 and Supplementary Table 9). To create this map, NPCs were immuno-purified from Mlp1-PPX-PrA *S. cerevisiae* strain, in a final buffer of 20 mM HEPES (at pH 7.5), 50 mM Potassium Acetate, 20 mM NaCl, 2 mM MgCl₂, 0.1% Tween 20 and 1 mM DTT (see ‘2. Immuno-purification of the endogenous *S. cerevisiae* NPC’ for details). The concentration was estimated by SDS-PAGE to be ~0.3–0.4 mg/ml. Freshly cleaned Quantafoil 300 mesh copper grids with 2-µm holes in the support film were prepared with a continuous carbon support film that spanned the holes. Before use, the grids were glow discharged in air, floated on 5 µl sample drops for 45 min and then washed by serial transfer on 4 × 20 µl drops of elution buffer without glycerol. Each grid was mounted on forceps in a Mark III Vitrobot (FEI) at room temperature and 100% relative humidity. Buffer on the grid was

removed by blotting from the bottom with a tool that held a filter paper wedge, using access through the left-hand port. Then, 2 µl of freezing buffer was added to the grid from the right-hand port and the grid was plunge-frozen in liquid ethane after blotting.

Cryo-ET data collection was done on a Titan-Krios electron microscope operating at 300 kV, equipped with an X-FEG, a post-column energy filter set to 20 eV and a spherical aberration (Cs) corrector (Supplementary Table 9). Images were recorded with a Gatan K2 Summit direct electron detector in integration mode, with single frames taken at each tilt with UCSF Tomo⁷⁰, at a nominal spacing of 5.6 Å per pixel. A total of 253 tilt series were collected in steps between –60°, 0° and 60° in increments of 2.5–4° for different tilt series. Although the full tilt range was used for tomogram reconstruction, in the final sub-tomogram averaging step only data up to ± 45° tilt from each sub-tomogram were included in the final average. The dose target for each tilt series was 90–100 electrons per Å² and followed a cosine α dose curve with a flux of 20 electrons per pixel per second, and a dose of 3.5 electrons per Å² for the zero tilt image. Extended Data Fig. 4a presents the strategy we used to reconstruct the 3D map of the whole yeast NPC. The 3dmod viewer in IMOD⁷¹ was used to visually screen tilts for defects with the Fourier transform, to gauge image motion. In total, 120 tilt series with a defocus range between –4.6 and –7.5 µm were kept for further processing (Supplementary Power Point Presentation slides 1, 2). After interactive test runs with etomo⁷¹, we processed the tilt series in an automated fashion with batchtomo using 7 × 7 patch tracking to create aligned tilt series and calculated tomograms with back-projection and the simultaneous iterative reconstruction technique (SIRT), which were contrast transfer function (CTF)-corrected by phase-flipping each image in the tilt series (Supplementary Power Point Presentation slides 3–6). The final SIRT tomograms were binned 3 × and used for interactive sub-tomogram ‘particle’ picking with e2spt_boxer.py in the EMAN2 single particle tomography package^{72,73} with a low pass filter of 100 Å. In total, 6,416 fully sampled unfiltered sub-tomograms were extracted from the back-projection tomograms in 300 × 300 × 300 voxel volumes.

In the alignment and averaging process, new algorithms for high-speed 3D alignment with automatic missing wedge compensation and averaging⁷³ were used throughout, and were critical for processing such large sub-tomograms. An initial reference was prepared by averaging a small subset of sub-tomograms to produce a low-resolution reference using the C₈-symmetry of the complex⁷⁴. Owing to the large size and distinct shape of the particles, alignments were unambiguous. The alignment and averaging strategy for the final map was adapted from previously described procedures⁷², and applied iteratively. The observed flexibility of the NPC ring initially limited the overall resolution to ~38 Å with 5,245 sub-tomograms (data not shown); the sub-tomograms discarded at this stage were those with the worst quality when compared to the average, generally owing to higher noise levels but in some cases due to particle damage or false positives during sub-tomogram picking. We realized that observed flexibility of the NPCs limited our resolution to ~38 Å and therefore used a tactic to locally align all individual spokes (C₈-symmetry units) to the reference rather than aligning whole rings, which could contain long-range deviations from a perfect toroid. It is important to note that these deviations are not large; across the entire NPC, they are on the order of the 38 Å resolution achieved without local alignment. All previous NPC cryo-ET maps have been produced using the approach of dividing the NPC into subunits^{6,29,75,76}. In brief, two reference volumes were prepared; (1) the entire NPC and (2) a masked volume in which one-fourth of the ring had been retained, centred roughly on the mass of a single subunit. Each NPC was rotationally and translationally aligned to the reference ring. Using this initial alignment, each NPC was replicated into its eight pseudo-symmetric orientations, then a translation-only alignment against the masked reference was performed. This had the effect of bringing one asymmetric unit per replicated ring into register with the reference at a consistent radius. Although small per-subunit rotations might have occurred, this possibility was not included in the local alignment. The average from the eight subunits was then used to construct a symmetric ring by applying an azimuthal linear ramp mask centred on the mask used for alignment, which fell to zero at an angle of 45° in both directions, and then imposing the C₈-symmetry. This interpolates smoothly from one side of the subunit to the other symmetry-related side, to produce a complete symmetrized ring. This processing dramatically reduces the blurring caused by local fluctuations in subunit position and the resulting 3D volume was used as a reference in the next cycle of iterative refinement, which was repeated until no further improvement was observed.

At this stage we realized that the preferred orientation of the particles within our tomograms was leading to anisotropic resolution in the final structure, with 2/3 of the NPC rings oriented within 30° of the C₈-symmetry axis (as clearly observed in our raw data). Producing an isotropic average required balancing the various ring orientations by discarding the lowest-contrast rings in the over-populated

orientations to more evenly balance the orientation distribution^{77–79}. This normalizing of orientations, and not 3D classification, was what led to discarding a fraction of our data (Extended Data Fig. 4a). When doing this, we elected to discard noisier sub-tomograms. This was achieved by comparing the agreement of each sub-tomogram with the overall average. In each angular range, we then retained roughly the same number of sub-tomograms, keeping those with the best quality. The discarded sub-tomograms were nearly as good, meaning we could equally well have used the next best subset of the sub-tomograms with virtually no effect on the final structure. Indeed, in the less common orientations, we were forced to use virtually all of the sub-tomograms irrespective of quality. Thus, in our work, a large fraction of the sub-tomograms were discarded not owing to their poor quality or conformational variability, but rather because of the preferred orientation of the particles within the tomogram. The final reconstruction used 1,864 (of the 6,416 initial) sub-tomograms. These sub-tomograms were further divided randomly into two groups for resolution assessment. ‘Gold standard’ refinement was used for resolution testing and to ensure self-consistency (Supplementary Power Point Presentation slides 7, 8). Both global and local resolution assessments were done using a set of tiled Gaussian masks to estimate the local resolution and reproducibility of the structure, which is one of the standard methods for local resolution assessment. In brief, a 3D Gaussian shape is generated with a ‘full width at half maximum’ parameter of at least $2\times$ the anticipated resolution, and generally even larger. This Gaussian shape is then applied as a mask to both maps and a Fourier shell correlation (FSC) curve is computed. This process is repeated in a tiled pattern throughout the volume. This provides a resolution for each sampled location in the volume. This procedure involves a trade-off: with smaller Gaussians, FSC curves are less precise, but with larger Gaussians the resolution estimate is less localized. The Gaussians overlap to provide better sampling of the volume. The resulting resolution map is similar to those produced by ResMap⁸⁰, but it measures FSC, which is filter-independent, unlike ResMap⁸⁰ that requires unfiltered volumes and would not work well on sub-tomogram averages. The global resolution of our cryo-ET map at the standard FSC_{0.143} cutoff is ~ 28 Å and the local resolution distribution ranges from 20 to 38 Å, with the inner ring being in the 20–25 Å range (Extended Data Fig. 5a–d). This local resolution estimate was used to locally filter the density map, to produce a map with the appropriate level of detail in each area (Extended Data Fig. 5). The size of the Gaussian window was 140 Å, indicating the smallest region over which the resolution is considered to vary. Although this may seem large, it is small compared to the size of the overall NPC (Extended Data Fig. 5). CTF phase-flipping was applied during tomographic reconstruction and a final approximate amplitude correction was applied to the averaged NPC ring. Therefore, theoretical CTF curves for the mean defocus values present in the tomograms were averaged assuming 10% amplitude contrast. The reciprocal of this curve was then applied as a filter to the final uncorrected map. The cryo-ET density map was refined at 5.3 Å per pixel on the basis of a recalibration of the map with known structures.

In parallel, and as an additional validation of our final map, we also carried out a tomographic analysis of the yeast NPC dataset (the same 6,416 sub-tomograms) using RELION 1.4, and incorporated a CTF model^{81,82}. In brief, we calculated back-projection tomograms without phase-flipping corrections for the individual tilted images, and binned the output sub-tomograms twofold to 10.6 Å per pixel. The datasets underwent sequential rounds of 2D classification using Z-projections of the sub-tomograms to eliminate poor particles. A subsequent 2D classification identified near-top, tilted and side views; the latter provided an independent estimate of NPC thickness perpendicular to the nuclear envelope (620–640 Å). A 3D reconstruction using the best sub-tomograms with RELION produced a map at ~ 35 Å resolution (data not shown) with similar features to those obtained with the e2spt in the EMAN2 single particle tomography package^{72,73}, including distinct connections between each spoke and the transporter, further validating these features in our cryo-ET map (Fig. 3). Finally, Z-projections of original sub-tomograms that were roughly aligned along the C₈-symmetry axis were used for an additional unsupervised 2D classification, which produced classes with central transporters without using the C₈-symmetry restraint (Extended Data Fig. 6a, b). Differences in the apparent resolution of the class averages in Extended Data Fig. 6a, b reflect different particle numbers in the classes. As mentioned above, the dataset of particles has a strong orientational bias, in which the NPCs tend to bind to the carbon support film with a range of 0–30 degrees of tilt. The class averages are based on 2D projections along the z axis of the original sub-tomograms, to avoid issues with the missing wedge, and there is therefore a disparity in particle numbers in the classes. Tilting in the tomographic data collection helped to fill in the missing data, but we took great care to ensure an equal coverage of Fourier space in the calculation of our final map, to avoid distortions, and also took a number of other steps to ensure that radiation damage and loss of data quality in later tilts was minimized by using only information in Fourier space at $\pm 45^\circ$ from each particle sub-tomogram

when they were combined to form the final map. The RELION map serves as a strong validation of our final map, because if our map was flawed, a reconstruction with RELION would have resulted in a different map (Extended Data Fig. 5e). Additionally, the fact that a RELION reconstruction resulted in a ~ 35 Å resolution map—virtually the same resolution as obtained in our ‘intermediate’ map described above (~ 38 Å)—validates our methodology and the quality of our final map (Extended Data Fig. 5e). An additional point that provides *prima facie* evidence that our cryo-ET map was calculated correctly is that local two-fold symmetry (C₂-symmetry), which was expected in the inner ring of the NPC, emerges without any enforcement, whereas the overall map shows a clear asymmetry with large and distinctly different features on the nuclear and cytoplasmic face of the yeast NPC (which were also observed in the RELION map) and a slightly tapered appearance, as is shown in Fig. 3a and Extended Data Figs 4d, 5b.

6. Small angle X-ray scattering. Small angle X-ray scattering (SAXS) measurements for 147 constructs of 18 Nups^{9,12,23,83–86} (Supplementary Table 6; S.J.K. *et al.*, manuscript in preparation; and Source Data) were carried out both at the Stanford Synchrotron Radiation Lightsources Beamline 4-2 in the SLAC National Accelerator Laboratory and at the SIBYLS Beamline 12.3.1 of the Advanced Light Source in the Lawrence Berkeley National Laboratory. SAXS data were collected at concentrations ranging from 0.5 to 5.0 (or higher, depending on the sample) mg/ml, using the previously defined standard protocol^{12,23,83}; approximately 20 one-second exposures were used for each sample and buffers at 15 °C. Further details of the SAXS experiments have previously been published^{9,12,23,83–86}.

7. Phenotypic analysis by one-cell doubling evaluation by living arrays of yeast. Yeast growth phenotypes were quantified using the one-cell doubling evaluation by living arrays of yeast (ODELAY) assay, as previously described¹⁴. In brief, yeast was cultured in YPD medium in 96-well plates overnight. Cultures were diluted to an OD₆₀₀ of 0.09 and allowed to grow for 6 h at 30 °C. The cultures were then diluted again to an OD₆₀₀ of 0.02 and spotted onto YPD agarose medium. The resulting cultures were then observed using time-lapse microscopy for 48 h with 30 min intervals between images. All images were collected on Leica DMI6000 microscopes with a 10×0.3 NA lens using bright field microscopy. MATLAB scripts using the Micro-Manager interface controlled the image collection process⁸⁷. Six independent experiments were performed. The population growth rates were scored against each other using the following equation:

$$Z_{\text{mean}} = \frac{1}{n} \sum_i^n \frac{d_i - \mu_i}{\sigma_i}$$

in which d_i is the i th decile of doubling time of the query population, μ_i is the mean of the i th decile of the doubling time of the parent strain and σ_i is the standard deviation of the i th decile of the doubling time of the parent strain. The mean and standard deviation deciles were calculated from at least 4 separate populations containing at least 200–300 individuals. All calculations were performed using MATLAB scripts. Following Z-scoring of the populations, an additional weight was added to the scoring for truncation strains that occurred in haploid versus diploid strains of yeast.

8. Negative-stain electron microscopy of the native Nic96 complex. An affinity-captured and natively eluted sample of the endogenous Nic96 complex (composed of Nic96, Nsp1, Nup49 and Nup57) was applied to a glow-discharged grid and stained with 1% uranyl formate. Images were collected on a Philips CM200 transmission electron microscope (FEI) operating at 200 kV at $50,000\times$ magnification and a defocus of ~ 1.5 μm (2.03 pixels per Å). Images were recorded on a Gatan UltraScan 1000 2k \times 2k CCD camera (Gatan). Particles were selected using Boxer from EMAN⁸⁸, normalized and then phase-flipped using ctfilt from EMAN. In total, 34 class averages (selected classes shown in Extended Data Fig. 7g) were generated through ISAC⁸⁹ that classified $\sim 86\%$ of the original set of 5,458 particles.

9. Integrative structure determination of the *S. cerevisiae* NPC. The structure of the *S. cerevisiae* NPC, including the scaffold, membrane rings, cytoplasmic export platform and nuclear baskets in the context of the pore membrane—but excluding the flexible FG regions—was solved by integrative structure determination (see ‘9.1 Integrative structure determination of the *S. cerevisiae* NPC scaffold, membrane rings, cytoplasmic export platform and nuclear basket’). Moreover, the distributions of the FG regions and the cargo-bound NTFs, comprising the central transporter, were computed by Brownian dynamics simulation (see ‘9.2 Brownian dynamics simulation of FG repeats and NTFs’).

9.1 Integrative structure determination of the *S. cerevisiae* NPC scaffold, membrane rings, cytoplasmic export platform and nuclear basket. Integrative structure determination of the *S. cerevisiae* NPC proceeded through four stages^{8,90–92} (Extended Data Fig. 1, Supplementary Table 3 and Supplementary Videos 1–3): (1) gathering data, (2) representing subunits and translating data

into spatial restraints, (3) configurational sampling to produce an ensemble of structures that satisfies the restraints and (4) analysing and validating the ensemble structures and data (Extended Data Figs 1, 7, 8 and Supplementary Tables 2–4). The integrative structure modelling protocol (stages 2, 3 and 4) was scripted using the Python modelling interface (PMI) package version 4d97507, which is a library for modelling macromolecular complexes based on our open-source integrative modelling platform (IMP) package⁹⁰ version 2.6 (<https://integrativemodelling.org>). The current procedure is an updated version of previously described protocols^{9,10,12,93–96}.

9.1.1 Stage 1: gathering data. The stoichiometry of Nups in the NPC was determined using native mass spectrometry and biochemical quantification of the purified NPC complex (Fig. 1 and Extended Data Figs 2, 3). In total, 3,077 intra- and intermolecular DSS and EDC unique cross-links were identified using mass spectrometry (Fig. 2 and Supplementary Table 1), which informed the spatial proximities among the 32 Nups and their conformations. The density map of the entire NPC was determined by cryo-ET at an average resolution of 28 Å, with the local resolution as high as ~20 Å for the inner ring, which informed the shape of the NPC (Fig. 3 and Extended Data Figs 4–6). Re-interpreted immuno-electron microscopy data^{3,8} informed the positions of 29 Nups. Predictions of the transmembrane domains obtained from the *Saccharomyces* Genome Database⁹⁷ (<http://yeastgenome.org>) and predictions of MBMs from the HeliQuest webserver^{12,98} informed about their respective proximities to the pore membrane. Previous immuno-electron microscopy measurements⁹⁹ informed the end-to-end distance for Mlp1 and Mlp2. Low-resolution electron microscopy images of the NPC⁴⁴ informed the diameter of the distal basket ring formed by Mlp1 and Mlp2.

Representations of individual Nups and some of their sub-complexes (Supplementary Table 2 and references therein) relied on (1) atomic structures of 21 yeast Nup domains and 3 sub-complexes determined by X-ray crystallography or nuclear magnetic resonance spectroscopy; (2) our previously determined structures of Nup116, Nup133, Nup145N, Nup192 and Pom152, as well as the Nup82 and Nup84 sub-complexes solved by integrative structure determination^{9,10,12,23,83–86}; (3) 29 comparative models built with MODELLER 9.13¹⁰⁰ on the basis of known structure(s) detected by HHpred^{101,102}; (4) SAXS profiles for 147 constructs of 18 Nups^{9,12,23,83–86} (Supplementary Table 6; S.J.K. *et al.*, manuscript in preparation); (5) secondary structure, disordered regions, and domain boundaries predicted by PSIPRED^{103,104}, DISOPRED¹⁰⁵, and DomPred¹⁰⁶, respectively; (6) coiled-coil regions of Nup82, Nup159, Nsp1, Nup49, Nup57, Mlp1, and Mlp2 predicted by COILS/PCOILS¹⁰⁷ and Multicoil2¹⁰⁸; (7) an atomic structure of the Nup53^{229–365} RRM domain from *S. cerevisiae* determined by X-ray crystallography (P. Sampathkumar *et al.*, manuscript in preparation); and (8) negative-stain electron microscopy density maps of full-length Nup192 (EMD-5556⁸⁶) and Pom152 (EMD-8543²³). See Supplementary Table 2 and references therein for all above (1) to (8).

Our previously published topological map of the NPC³ and the 82 composites determined by affinity purification and overlay assay⁸ were not used for computing the current NPC structure, but were used for validating the current NPC structure.

9.1.2 Stage 2: representing subunits and translating data into spatial restraints. Information about the modelled system (see ‘9.1.1 Stage 1: gathering data’) can in general be used for defining the system’s representation, defining the scoring function that guides sampling of alternative structural models, limiting sampling, filtering of good-scoring structures obtained by sampling and final validation of the structures. Here the NPC representation relies primarily on stoichiometry as well as atomic structures, integrative structures, comparative models and SAXS profiles of Nups and their sub-complexes (Supplementary Tables 2 and 6, and references therein); the scoring function relies on chemical cross-links, the cryo-ET density map, immuno-electron microscopy localizations, excluded volume, sequence connectivity, the shape of the pore membrane and four types of sequence-based localization relative to the membrane (below); the sampling benefits from symmetry constraints (below); and the validation of the final structure relies in part on the SAXS profiles (Supplementary Table 6) and composites determined by affinity purification and overlay assays⁸ (below).

To improve computational efficiency and avoid a representation that was too coarse, we represented the NPC in a multi-scale fashion. A rigid-body consisting of multiple beads was defined for each X-ray structure, NMR structure, comparative model and integrative structure of the NPC components (Supplementary Table 2). The remainders of the Nup sequences not in rigid-bodies (36.8% of residues, excluding FG repeats) were represented as flexible strings of beads. In a rigid-body, the beads have their relative distances constrained during configurational sampling, whereas in a flexible string the beads are restrained by the sequence connectivity, excluded volume and potentially additional restraints, such as chemical cross-links, as exemplified in previous publications^{9,10,23,93,109}.

Rigid-bodies (63.2% of residues, excluding FG repeats) were coarse-grained using two resolutions, in which beads represented either individual residues or segments of up to ten residues. The coordinates of a 1-residue bead were those of the corresponding C α atom. The coordinates of a 10-residue bead were the centre of mass of the ten constituent 1-residue beads. Finally, the remaining regions without an atomic representation (that is, the predicted transmembrane and disordered regions) were represented by a flexible string of beads encompassing 25 to 100 residues each; the low-resolution representation of these regions is justified because their conformations are likely to be ‘decoupled’ from the structure of the rest of the NPC^{3,110}.

We used the SAXS data to confirm the rigid-body representations of eight Nups with X-ray structures, comparative models and previously published atomic integrative structures^{9,12,23,83–86} (Extended Data Fig. 7f and Supplementary Tables 2, 6). The rigid-body representation of a Nup construct was validated by a χ -value that quantifies the difference between the computed (from an atomic rigid-body representation using FoXS¹¹¹) and experimental SAXS profiles, except for several constructs of Nup133 and Nup192 that were flexible during integrative modelling and were thus evaluated as previously described^{12,86,112}. The χ -value validation assumes that each Nup construct, corresponding in most cases to a single domain (not the whole protein), has the same conformation in solution and in complex; this assumption is consistent with other data (for example, the chemical cross-links and cryo-ET map). The SAXS validation is necessarily limited to *S. cerevisiae* constructs of Nups that exist as a rigid monomer in solution and do not contain FG repeats; rigid-body representations of the constructs from other species, constructs that oligomerize in solution and constructs that include FG repeats cannot be easily used for validation, because of the sensitivity of a computed SAXS profile to the differences in the sequence and stoichiometry, as well as to potential errors in comparative modelling (especially of insertion and deletion).

After producing this validated representation, we next encoded the spatial restraints on the basis of information gathered in Stage 1, according to the following steps (Supplementary Table 4; for the definition of the scoring function consisting of these restraints, see ‘9.1.3.2 Scoring function’):

(1) Cross-link restraints: 1,643 of the 3,077 unique cross-links (Fig. 2 and Supplementary Table 1a) were used to restrain the distances spanned by the cross-linked residues, relying on a Bayesian scoring function¹⁰. The evaluation takes into account the ambiguity due to multiple copies of identical subunits and, for cross-links involving the same protein type, due to the lack of knowledge of whether they are intra- or intermolecular^{9,93,109}, the ambiguous cross-link restraint considers all intra- and intermolecular assignments in multiple copies of identical subunits, with only the least violated distance contributing to the score. The remaining 1,434 DSS and EDC cross-links (Fig. 2 and Supplementary Table 1b–f) were already used as restraints to build the integrative structures of the Nup84¹⁰ and Nup82 sub-complexes⁹, represented here as rigid-bodies. The two homo-dimer DSS cross-links between two copies of residue 62 of Pom152²³ and two copies of residue 151 in Nup60²² were transformed into harmonic upper-distance bounds, enforcing the homo-dimer configuration.

(2) Cryo-ET density restraint: the cryo-ET density restraint was applied, which corresponded to the cross-correlation between the Gaussian mixture model (GMM) representation of most Nups and the GMM representation of the cryo-ET density map^{95,113–115} (Fig. 3 and Extended Data Figs 4–6); we used a GMM representation for the sake of computational efficiency, necessitated by the large size of the NPC. An assessment of a given structure against a density map is much faster when both are represented with a mixture model (because the number of components in a mixture model is much smaller than the number of grid points covering the maps). However, these two scores are very strongly correlated. Thus, the structures obtained with a grid representation, if we had sufficient computational power, would certainly be indistinguishable from the current NPC structures¹¹⁵.

A 90° arc of the cryo-ET density map was approximated by the GMM, which contained 1,750 components computed using the expectation-maximization algorithm implemented in scikit-learn (<http://scikit-learn.org>); the cryo-ET GMM appeared to be sufficient to reproduce the complete features of the density map (excluding the central transporter region). To use a comparable number of GMM components for Nups, a Nup was approximated by a GMM component for each of its 100 to 500 residues. The cross-correlation quantified the degree of overlap between the Nup GMM components and the cryo-ET GMM components.

(3) Immuno-electron microscopy localization restraints: the immuno-electron microscopy localization restraint was used to localize the C-terminal beads of 29 of the 32 Nups, on the basis of previous immuno-electron microscopy data^{3,8,116}. This goal was achieved by imposing upper and lower harmonic bounds on the axial and radial coordinates of the restrained bead, reflecting the uncertainty in the immuno-electron microscopy data⁸. The three remaining Nups (Nsp1, Sec13

and Seh1) were not restrained by the immuno-electron microscopy data because of their high uncertainty, presumably due to the positional heterogeneity of the tagged Nup in the multiple superposed electron microscopy images of the NPC. This heterogeneity is more likely to occur for Nups with multiple copies per C_2 -symmetry unit, which are unlikely to share the same radial and axial coordinates.

(4) Excluded volume restraints: the protein excluded volume restraints were applied to each 10-residue bead, using the statistical relationship between the volume and the number of residues that it covered^{8–10,117}.

(5) Sequence connectivity restraints: we applied sequence connectivity restraints, using a harmonic upper bound on the distance between consecutive beads in a subunit, with a threshold distance equal to three times the sum of the radii of the two connected beads. The bead radius was calculated from the excluded volume of the corresponding bead, assuming standard protein density^{8–10,117}.

(6) Membrane exclusion restraints: the membrane exclusion restraints were applied to beads in the non-membrane-spanning Nups or to their segments to prevent these beads from penetrating the pore membrane. A lower harmonic bound at 0 Å was applied to the distance between a bead and the closest point on the pore-side membrane surface^{3,8} (modelled as a half-torus with the large and small radii of 390 and 150 Å, respectively; Supplementary Table 3), for all coarse beads (10 residues or more per bead) in all Nups but Pom152, Ndc1 and Pom34; the restraint was also applied to all non-membrane coarse beads of Pom152^{1–110}, Ndc1^{1–28}, Ndc1^{248–655}, Pom34^{1–63} and Pom34^{151–299}.

(7) Transmembrane domain restraints: the transmembrane domain restraint was used to localize the coarse beads in the predicted transmembrane domains (Pom152^{111–200}, Ndc1^{29–247} and Pom34^{64–150}; Supplementary Table 2 and references therein) within the pore membrane, which is 45 Å thick^{3,8}. This aim was achieved by imposing an upper harmonic bound at 45 Å and a lower harmonic bound at 0 Å on the distance between the bead and the closest point on the pore-side membrane surface.

(8) Membrane surface binding restraints: the membrane surface binding restraint was used to localize the coarse beads in the predicted MBMs (Nup1^{1–32}, Nup60^{27–47}, Nup120^{135–152}, Nup120^{197–216}, Nup133^{252–270}, Nup157^{310–334}, Nup170^{320–344}, Nup53⁴⁷⁵ and Nup59⁵²⁸; Supplementary Table 2 and references therein), within the pore membrane up to 12 Å from the pore-side membrane surface¹¹⁸. This aim was achieved by imposing an upper harmonic bound at 12 Å within the pore membrane and a lower harmonic bound at 0 Å on the distance between the bead and the closest point on the pore-side membrane surface. For Nup120, only the best satisfied of the Nup120^{135–152} and Nup120^{197–216} restraints was used^{10,12} (conditional restraint).

(9) Pom152 perinuclear volume restraint: only the C-terminal region of Pom152 (residues 201–1337) was restrained to the perinuclear lumen of the pore membrane²³. This aim was achieved by imposing a lower harmonic bound at 0 Å on the distance between the Pom152 beads and the closest point on the perinuclear side of the membrane surface.

(10) Distal basket ring restraints: the conformations of Mlp1 and Mlp2 were restrained by an upper harmonic bound at 350 Å and a lower harmonic bound at 230 Å on the distance between the N-terminal and C-terminal beads, on the basis of immuno-electron microscopy measurements⁹⁹. In addition, the radius of the distal basket ring was restrained by an upper harmonic bound at 170 Å and a lower harmonic bound at 130 Å on the radial coordinates of the C-terminal beads of Mlp1 and Mlp2, on the basis of low-resolution electron microscopy images of the NPC⁴⁴. The nuclear basket was also informed by cross-linking restraints and the C_8 -symmetry constraint (see '9.1.3.1 Sampling space with symmetry constraints').

9.1.3 Stage 3: Configurational sampling. We used the configurational sampling to produce an ensemble of structures that satisfies the restraints, as described below **9.1.3.1 Sampling space with symmetry constraints.** We aimed to maximize the efficiency of the configurational sampling: more specifically, we aimed to maximize the precision at which the sampling of good-scoring solutions was exhaustive (see '9.1.4 Stage 4: analysing and validating the ensemble structures and data'). Therefore, we reduced the number of independently moving parts in the NPC structure by explicitly considering the C_8 - and C_2 -symmetries of the NPC, as follows. The entire NPC consists of 8 clones of the C_8 -symmetry unit, related by multiples of a 45° rotation around the z axis (Fig. 3 and Extended Data Figs 4–6). The C_8 -symmetry unit was further broken into two C_2 -symmetry units and non- C_2 -symmetric Nups (Supplementary Table 2a); the C_2 -symmetry unit contains Nups that occur equally on both the cytoplasmic and nucleoplasmic sides^{3,8,116}. For computational efficiency, we defined the coordinate system such that the C_2 -symmetry is imposed simply by cloning a bead in the C_2 -symmetry unit at (x, y, z) to $(x, -y, -z)$ (equivalent to a rotation of 180° around the x axis). This aim was achieved by fitting both copies of Pom152²³ into the cryo-ET density map, followed by moving the centre of the map to the origin of the coordinate

system and orienting the map such that the x , $-y$, $-z$ transformation applies to Pom152.

With these symmetries in hand, we sampled only the positions of rigid-bodies and beads corresponding to the Nups in the C_2 -symmetry unit and non- C_2 -symmetric Nups. There are no Nups that occur on both the cytoplasmic and nuclear sides and are not related by the C_2 -symmetry; there are no Nups that occur with a different stoichiometry on both sides. In addition, the luminal domain of Pom152 was considered already well-positioned given its fit into the cryo-ET density map (Fig. 3e) and peripheral location in the NPC²³, and was not sampled further.

9.1.3.2 Scoring function. The scoring function included restraints on the sampled Nups and the Pom152 luminal domain as well as restraints across the interfaces with neighbouring symmetry units: (1) the cryo-ET density restraint and distal basket ring restraint applied to the Nups in the sampled C_8 -symmetry unit; (2) sequence connectivity, immuno-electron microscopy localization, and the four types of sequence-based localizations relative to the membrane applied to the Nups in the sampled C_2 -symmetry unit and non- C_2 -symmetric Nups; and (3) cross-link and excluded volume restraints applied to the pairs of beads for Nups within the sampled C_8 -symmetry unit and across the interfaces with neighbouring symmetry units.

9.1.3.3 Sampling algorithm. The search for good-scoring structures relied on replica exchange Gibbs sampling, based on the Metropolis Monte Carlo algorithm^{9,10} (Supplementary Table 3). The Monte Carlo moves included random translation and rotation of rigid-bodies (up to 4 Å and 0.04 radians, respectively) and random translation of individual beads in the flexible segments (up to 4 Å). As indicated above, these operations were applied only to the sampled rigid-bodies and beads. The remaining, symmetry-constrained rigid-bodies and units were moved in lockstep to maintain the exact C_8 - and C_2 -symmetries at each sampling step, as described above. Up to 64 replicas were used, with a 1.0–5.0 temperature range. Forty-two independent sampling calculations were performed, each one starting with a random initial configuration. The coordinates were saved every 10 Gibbs sampling steps, each consisting of a cycle of Monte Carlo steps that moved every rigid-body and flexible bead once.

To further increase the efficiency of sampling, we first applied the above Monte Carlo algorithm separately to the following four subsets of Nups, which are co-localized on the basis of previous characterizations^{3,8,9,23} and the current cryo-ET density map: (1) Nup82 and Nup84 sub-complexes, (2) Pom152, (3) inner-ring Nups (Nup157, Nup170, Nup188, Nup192, Pom34, Ndc1, Nup53, Nup59 and Nic96^{205–839}), and (4) Mlp1 and Mlp2. Next, the best-scoring solutions from sampling each of the first three subsets were combined; they were already in the same reference frame, because they were all obtained by fitting the same cryo-ET density map and immuno-electron microscopy data. The rest of the Nups and the Mlp1–Mlp2 heterodimer were then added in random positions and orientations, followed by another application of the above Monte Carlo algorithm to all sampled Nups. This sampling produced a total of 100,453 modelled structures in 42 independent runs (the score ranges from 88,545.0 to 103,589.5, with the mean and standard deviation of 88,831.5 and 187.4, respectively), requiring ~10 weeks on a cluster of ~2,500 CPU cores. For the most detailed specification of the sampling procedure, see the IMP modelling script (<https://salilab.org/npc2018>).

We considered for further analysis only the 5,529 modelled structures with the scores better than 88,644.1 (1 standard deviation below the mean value); this threshold implies satisfaction of the input datasets within their uncertainties (Supplementary Table 4; see '9.1.4.3 Fit to input information'). These structures are already superposed because they were fit into the same Cryo-ET map and sampling did not move the luminal domain of Pom152 (see '9.1.3.1 Sampling space with symmetry constraints').

9.1.4 Stage 4, analysing and validating the ensemble structures and data. Input information and output structures need to be analysed to estimate structure precision and accuracy, detect inconsistent and missing information, and to suggest more informative future experiments. We used the previously published analysis and validation protocol^{8,9}. Assessment began with a test of the thoroughness of structural sampling, followed by structural clustering of the modelled structures and estimating their precision based on the variability in the ensemble of good-scoring structures, quantification of the structure fit to the input information and structure assessment by data not used to compute it; structure assessment by cross-validation was not performed in this case, because it takes ~10 weeks on approximately 2,500 CPU cores to compute an ensemble of structures for a single set of input datasets. These validations are based on the nascent wwPDB effort⁹² toward archiving, validating and disseminating integrative structures. We now discuss each one of these validations in turn.

9.1.4.1 Thoroughness of the configurational sampling. We must first estimate the precision at which sampling found the most good-scoring solutions (sampling precision); the sampling precision must be at least as high as the precision of the

structure ensemble that is consistent with the input data (structure precision). As a proxy for testing the thoroughness of sampling, we performed four tests of sampling convergence¹¹⁹, as follows.

The first convergence test confirmed that the scores of refined structures do not continue to improve as more structures are computed, essentially independently of each other (Extended Data Fig. 1c).

The second convergence test confirmed that the good-scoring structures in independent sampling runs 1–21 (structure sample 1; $n^{\text{sample1}} = 2,359$ structures) and 22–42 (structure sample 2; $n^{\text{sample2}} = 3,170$ structures) satisfied the data equally well. The non-parametric Kolmogorov–Smirnov two-sample test^{120,121} (two-sided) indicates that the difference between the two score distributions is insignificant (P value (1.0) > 0.05). In addition, the magnitude of the difference is small, as demonstrated by the Kolmogorov–Smirnov two-sample test statistic, D , of 0.045 (Extended Data Fig. 1d). Thus, the two score distributions are effectively equal.

Next, we considered the 5,529 good-scoring structures themselves (not their scores as in the two tests described above). For stochastic sampling methods, thoroughness of sampling can be assessed by showing that multiple independent runs (for example, using random starting configurations and different random number generator seeds, as is the case for structure samples 1 and 2) do not result in noticeably different structures^{8–10,13}. We tested the similarity between structure samples 1 and 2 in the following two ways.

The third convergence test¹¹⁹ relied on the χ^2 -test (one-sided) for homogeneity of proportions¹²² between structure samples 1 and 2 (Extended Data Fig. 1e, f). The test involves clustering structures from both samples, followed by comparing the proportions of structures from each sample in each cluster. No adjustment was made for multiple comparisons. A comparison of two NPC structures considered only the beads representing Nups with a single copy per C_2 -symmetry unit and the Nic96 complex (including all Nups in the inner, outer and membrane rings, but excluding Nup100, Nup116, Nup145N, Nup1, Nup60, Gle1, Nup42, Mlp1 and Mlp2), to avoid the combinatorial explosion in identification of topologically equivalent Nup copies. The sampling precision is defined as the largest root-mean-square deviation (r.m.s.d.) between a pair of NPC structures within any cluster, in the finest clustering for which each sample contributes structures proportionally to its size (considering both the significance and magnitude of the difference) and for which a sufficient proportion of all structures occur in sufficiently large clusters. The sampling precision for our NPC structure is 9 Å (Extended Data Fig. 1e).

Threshold-based clustering¹²³ results in a single dominant cluster containing 80.3% of the good-scoring structures (Extended Data Fig. 1e, f) with a root-mean-square fluctuation (r.m.s.f.) of 9 Å (cluster precision). The remaining 19.7% of the structures are similar to those in the dominant cluster; the largest r.m.s.d. value from a structure in the dominant cluster is 17 Å (the mean and standard deviation of the r.m.s.d. values are 13.3 and 1.3 Å, respectively). Therefore, there is effectively a single good-scoring solution, at the structure precision of 9 Å (equal to the cluster precision). The sampling precision of 9 Å (r.m.s.d.) is sufficiently high for computing a structure at 9 Å precision (r.m.s.f.; r.m.s.d. is approximately $\sqrt{2} \times \text{r.m.s.f.}$ ¹²⁴). For the remainder of our analysis, we use only the structures in the dominant cluster.

The fourth convergence test relied on a comparison of two localization probability density maps for each Nup, obtained for dominant cluster structures in samples 1 and 2. A localization probability density map defines the probability of any voxel (here, $6 \times 6 \times 6 \text{ Å}^3$) being occupied by a specific protein in a set of structure densities, which in turn are obtained by convolving superposed structures with a Gaussian kernel (here, with a standard deviation of 5.4 Å). The average cross-correlation coefficient between the two maps for each Nup is 0.90, indicating that the positions of most Nups in the two samples are nearly identical at the structure precision of 9 Å.

In conclusion, all four sampling tests indicate that the sampling was exhaustive at 9 Å precision (Supplementary Table 3). The caveat is that passing these tests is necessary but not sufficient evidence of thorough sampling; a positive outcome of the tests may be misleading if, for example, the landscape contains only a narrow—and thus difficult to find—pathway to the pronounced minimum corresponding to the correct structure. Moreover, our sampling was not completely stochastic because it proceeded in two steps, the first of which prepared the starting configuration for the second step. As a result, the actual structure precision might be worse^{125–128} than the estimated 9 Å.

9.1.4.2 Clustering and structure precision. An ensemble of good-scoring structures needs to be analysed in terms of the precision of its structural features^{3,8,9}. The precision of a component position can be quantified by its variation in an ensemble of superposed good-scoring structures. It can also be visualized by the localization probability density for each of the components of the NPC structure.

As described above, integrative structure determination of the NPC resulted in effectively a single good-scoring solution, at the precision of $\sim 9 \text{ Å}$. This precision is sufficiently high to pinpoint the locations and orientations of the constituent Nups, demonstrating the quality of the input data, including the chemical cross-links (Fig. 2 and Extended Data Fig. 7a–c) and the cryo-ET density map (Fig. 3 and Extended Data Fig. 8).

9.1.4.3 Fit to input information. An accurate structure needs to satisfy the input information used to compute it. Because the sampling was exhaustive at $\sim 9 \text{ Å}$ precision, overfitting is not a problem at this precision; all structures at this precision that are consistent with the data are provided in the ensemble.

The dominant cluster satisfies 90% of the DSS cross-links (Extended Data Fig. 7a–c and Supplementary Tables 1, 4); a cross-link restraint is satisfied by a cluster of structures if the corresponding C_α – C_α distance in any of the structures in the cluster (considering restraint ambiguity) is $< 35 \text{ Å}$ (Extended Data Fig. 7a–c; shown in blue). Therefore, the dominant cluster essentially satisfies the cross-linking data within its uncertainty (the false detection rate is approximately 5% to 10%^{129,130}). Most of the cross-link violations are small, and can be rationalized by local structural fluctuations, coarse-grained representations of some Nup domains, and/or finite structural sampling, as shown in Extended Data Fig. 7a (a histogram presenting the distribution of the cross-linked C_α – C_α distances).

The localization probability densities for the dominant cluster overlap well with the cryo-ET density map, with the cross-correlation coefficient of 0.92 (Fig. 3, Extended Data Fig. 8 and Supplementary Table 4). Additional density is present in the cryo-ET map for the Nup82 complex (cytoplasm) and basket attachment sites (nucleoplasm). This density may arise from local flexibility of these modules or may be due to the presence of cargo or transport factors associated with the NPC (Fig. 1b, c and Extended Data Fig. 3c). For visualization, the localization probability densities are typically smoothed and contoured at the threshold that results in approximately twice the protein volume estimated from its sequence (Fig. 4).

The remainder of the restraints are harmonic, with a specified standard deviation. The dominant cluster generally satisfied at least 95% of restraints of each type (Supplementary Table 4); a restraint is satisfied by a cluster of structures if the restrained distance in any structure in the cluster (considering restraint ambiguity) is violated by less than 3 standard deviations, specified for the restraint. Most of the violations are small, and can be rationalized by local structural fluctuations, coarse-grained representations of some Nup domains and/or finite structural sampling.

9.1.4.4 Satisfaction of data and considerations that were not used to compute structures. The most direct test of a modelled structure is by comparing it to the data that were not used to compute it (a generalization of cross-validation).

First, our current NPC structure is consistent with our previously published data and topological map^{3,8} (Extended Data Fig. 7d, e). Our current structure satisfies all 82 composites determined by affinity purification and overlay assays^{3,8}, even though these were not used in this calculation. For example, Pom152, Pom34, Ndc1, Nup157 and Nup170 are connected with each other (left panel in Extended Data Fig. 7e), consistent with the composites determined in a previous publication using the affinity purification data^{3,8} (right panel in Extended Data Fig. 7e). Moreover, the position of each Nup in the current structure is generally similar to that in the previous topological map^{3,8}, although the current structure is determined at a precision that is an order of magnitude higher than in the previous map (Extended Data Fig. 7d).

Second, the atomic structures of eight Nups are consistent with the corresponding SAXS profiles for their constructs (Extended Data Fig. 7f and Supplementary Tables 2, 6), as discussed in ‘9.1.2 Stage 2: representing subunits and translating data into spatial restraints’. For example, the SAXS profile calculated from the atomic structure of Pom152^{718–1148} (red curve in Extended Data Fig. 7f) using FoXS¹¹¹ is well matched ($\chi = 1.48$) to the corresponding experimental SAXS profile²³ (black dots in Extended Data Fig. 7f; $n = 20$ exposures). For visualization purposes, the Pom152^{718–1148} structure (represented as a ribbon) is shown along with the best fit of the *ab initio* shape (represented as a transparent envelope) computed from the experimental SAXS profile, in Extended Data Fig. 7f.

Third, the structures of the Nic96 complex (composed of Nic96, Nsp1, Nup49 and Nup57) in the dominant cluster can be projected well on most of the 2D class averages obtained for the natively isolated complex (Extended Data Fig. 7g; see ‘8. Negative-stain electron microscopy of the native Nic96 complex’). More specifically, the electron microscopy 2D validation fits the structure of the Nic96 complex in the whole NPC context to the electron microscopy class averages of the Nic96 complex, and computes a score that quantifies the match. The computation proceeds in three stages: (1) generation of alternative model projections, (2) alignment of the class average and each model projection, and (3) calculation of the fitting score for each projection, as follows. First,

1,000 uniformly distributed projections of the low-pass-filtered structure of the NPC on the sphere (stage 1) were generated. Second, each projection was optimally aligned to each of the class averages in Fourier space (stage 2). Finally, a score corresponding to the cross-correlation coefficient was computed (stage 3). For example, the experimental class averages were satisfied by the structure with cross-correlation coefficients of 0.85 and 0.80, respectively (Extended Data Fig. 7g).

Fourth, the structure was also validated by its comparison to the core scaffold maps of the *Homo sapiens* NPC, which are based primarily on electron microscopy density maps^{6,7,17,29} (Extended Data Fig. 12). Overall, the inner-ring architecture is similar in both yeast and vertebrates, consistent with it being the most conserved part of the NPC³⁹.

Finally, the structure allows us to rationalize the functional fitness (Fig. 5), the transport through the NPC (Fig. 6 and Extended Data Fig. 11) and the evolution (Extended Data Fig. 10), therefore increasing our confidence in the structure compared to not being able to rationalize these aspects of the NPC^{8,131,132}.

9.2 Brownian dynamics simulation of FG repeats and NTFs. Distributions of the FG repeats and NTFs were computed by Brownian dynamics simulation¹³³, using our previously published protocol³⁴ implemented in IMP⁹⁰ version 2.6. The simulated system included the static NPC ring determined in this study, the pore membrane, disordered and flexible FG-repeat domains as well as freely diffusing NTFs and inert macromolecules, all enclosed within a bounding box of $2,000 \times 2,000 \times 2,000 \text{ \AA}^3$.

The pore membrane was represented as a 250 \AA slab with a cylindrical pore of radius 375 \AA that contains the static NPC ring (this pore membrane representation is simplified compared to the toroidal pore used for solving the structure of the static NPC ring, for reasons of computational efficiency). Each of the FG-repeat domains was represented as a flexible string of beads; a bead had a radius of 6 \AA and encompassed 20 residues to achieve a compromise between computational efficiency and accuracy^{34,110,134–136}. Consecutive beads were restrained by a bond with an equilibrium length of 18 \AA and a constant force of $1.0 \text{ kcal per mol per \AA}$, approximating the spring-like nature of flexible polymers¹³⁷ in general and FG-repeat domains^{110,134–136,138–140} in particular. The freely diffusing molecules included 1,600 NTFs and 1,600 inert macromolecules (0.33 mM each), each consisting of 8 subgroups of 200 macromolecules, ranging in radius from 4 to 28 \AA in 2 \AA increments (10 to 75 kDa , assuming constant protein density of 1.38 g/cm^3). Excluded volume interactions were applied to pairs of overlapping beads and to beads penetrating the pore membrane or the bounding box, using a constant repulsive force of $10.0 \text{ kcal per mol per \AA}$. The potential binding energy between a binding site on an FG motif and a binding site on an NTR was modelled by an anisotropic harmonic potential dependent on the distance and orientation between the two sites that reproduces measured apparent dissociation constants in our simulations (B.R. *et al.*, manuscript in preparation).

The Brownian dynamics of the entire system were simulated at 297.15 K for 40 microseconds with a time step of $1,047$ femtoseconds, independently 400 times; the first 10 microseconds of each trajectory were considered equilibration time and ignored in subsequent analysis. The distributions of the FG Nup and NTR positions were then computed from the total of 12 milliseconds of simulations over a cubic grid with a voxel size of $10 \times 10 \times 10 \text{ \AA}^3$, at time intervals of 9.5 picoseconds, from all 400 trajectories; these distributions were averaged by relying on the C_8 -symmetry of the NPC.

Code availability. Files containing integrative structure modelling scripts, as well as the input data and output results are available at <http://salilab.org/npc2018>. The source code for differential imaging is available at https://github.com/jayunruh/Jay_Plugins.

Data availability. Source Data for Fig. 1a are provided as an excel file. Original data underlying the image calibration data (source data for Fig. 1 and Extended Data Fig. 3) can be accessed from the Stowers Original Data Repository at <http://www.stowers.org/publications/LIBPB-1267>.

Raw data for the chemical cross-links (source data for Fig. 2 and Supplementary Table 1) are available via the Zenodo data repository, at <http://doi.org/10.5281/zenodo.1149746>.

The cryo-ET density map (source data for Fig. 3) is deposited in the Electron Microscopy Data Bank (EMDB) with the accession code EMD-7321.

The cryo-ET raw data (120 tilt series; source data for Extended Data Figs 4–6) are deposited in the Electron Microscopy Public Image Archive (EMPIAR) (<https://www.ebi.ac.uk/pdbe/emdb/empiar/>) with the accession code EMPIAR-10155 (see Supplementary Power Point Presentation slides 1, 2 as examples).

The integrative NPC structure (source data for Fig. 4) is deposited in the nascent public Protein Data Bank (PDB) repository, PDB-dev (<https://pdb-dev.wwpdb.org/>), under the accession codes PDBDEV_00000010, PDBDEV_00000011 and PDBDEV_00000012.

Source data for Extended Data Fig. 2b are provided in the Supplementary Information (Supplementary Fig. 1).

SAXS data (source data for Extended Data Fig. 7f) are deposited in the Small Angle Scattering Biological Data Bank (SASBDB; <https://www.sasbdb.org/>), under the accession codes SASDBV9, SASDBW9, SASDBX9, SASDBY9 and SASDBZ9. In addition, all SAXS data (Supplementary Table 6) are provided as source data with the article.

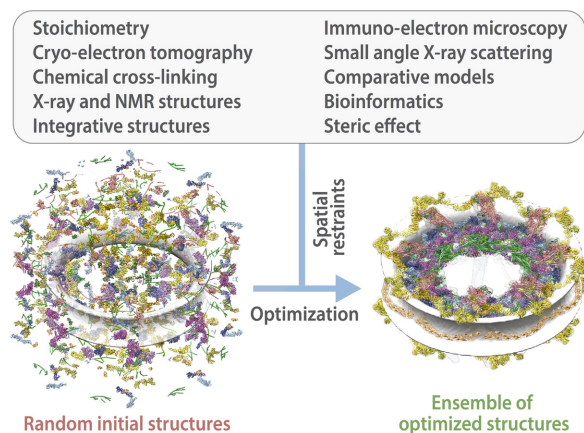
Raw data for the negative-stain electron microscopy of the native Nic96 complex (source data for Extended Data Fig. 7g) are deposited in the Electron Microscopy Public Image Archive (EMPIAR) (<https://www.ebi.ac.uk/pdbe/emdb/empiar/>) with the accession code EMPIAR-10162.

All other data are available from the corresponding author upon reasonable request.

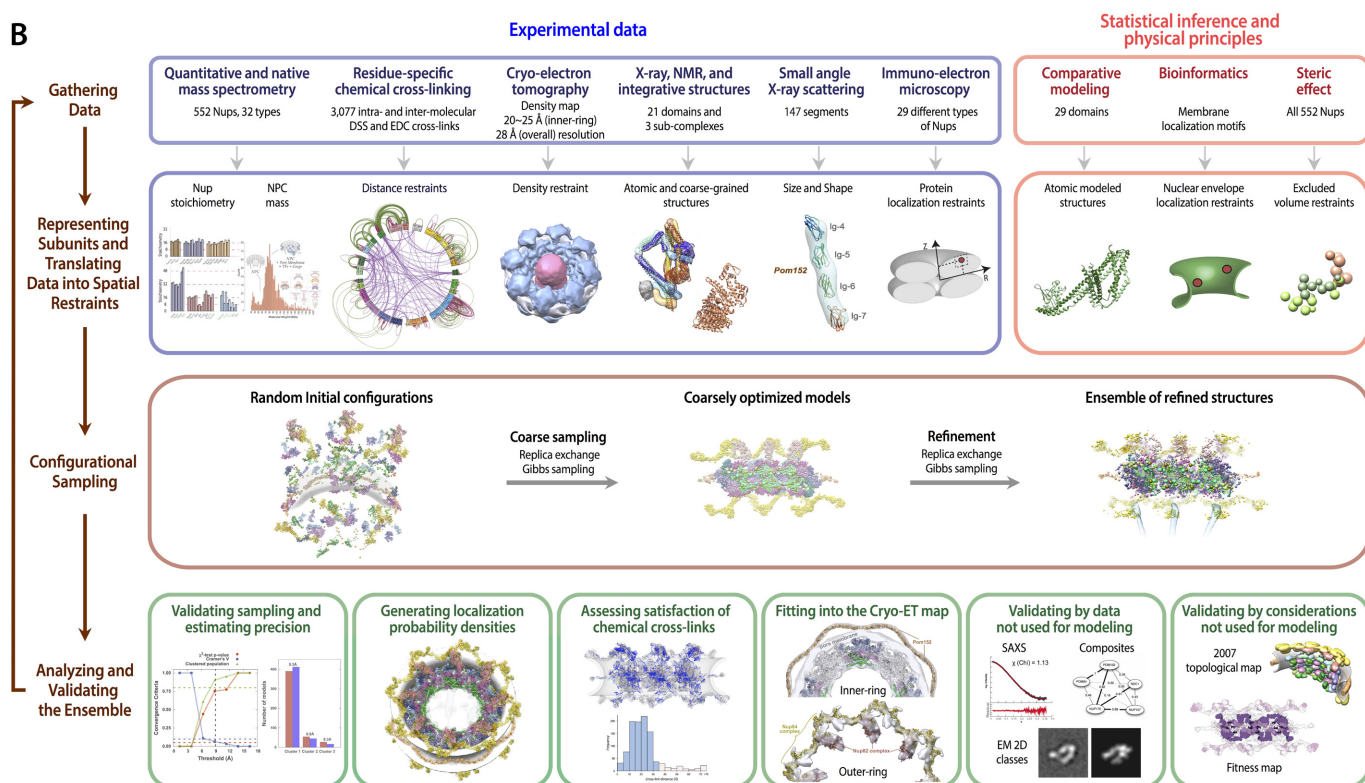
44. Rout, M. P. & Blobel, G. Isolation of the yeast nuclear pore complex. *J. Cell Biol.* **123**, 771–783 (1993).
45. LaCava, J., Fernandez-Martinez, J., Hakhverdyan, Z. & Rout, M. P. Protein complex purification by affinity capture. *Cold Spring Harb. Protoc.* <http://doi.org/10.1101/pdb.top077545> (2016).
46. LaCava, J., Fernandez-Martinez, J., Hakhverdyan, Z. & Rout, M. P. Optimized affinity capture of yeast protein complexes. *Cold Spring Harb. Protoc.* <http://doi.org/10.1101/pdb.top087932> (2016).
47. LaCava, J., Fernandez-Martinez, J. & Rout, M. P. Native elution of yeast protein complexes obtained by affinity capture. *Cold Spring Harb. Protoc.* <http://doi.org/10.1101/pdb.top087940> (2016).
48. Oeffinger, M. *et al.* Comprehensive analysis of diverse ribonucleoprotein complexes. *Nat. Methods* **4**, 951–956 (2007).
49. Hakhverdyan, Z. *et al.* Rapid, optimized interatomic screening. *Nat. Methods* **12**, 553–560 (2015).
50. Beynon, R. J., Doherty, M. K., Pratt, J. M. & Gaskell, S. J. Multiplexed absolute quantification in proteomics using artificial QCAT proteins of concatenated signature peptides. *Nat. Methods* **2**, 587–589 (2005).
51. Shivaraju, M. *et al.* Cell-cycle-coupled structural oscillation of centromeric nucleosomes in yeast. *Cell* **150**, 304–316 (2012).
52. Keifer, D. Z., Motwani, T., Teschke, C. M. & Jarrold, M. F. Measurement of the accurate mass of a 50 MDa infectious virus. *Rapid Commun. Mass Spectrom.* **30**, 1957–1962 (2016).
53. Pratt, J. M. *et al.* Multiplexed absolute quantification for proteomics using concatenated signature peptides encoded by QconCAT genes. *Nat. Protocols* **1**, 1029–1043 (2006).
54. Kito, K., Ota, K., Fujita, T. & Ito, T. A synthetic protein approach toward accurate mass spectrometric quantification of component stoichiometry of multiprotein complexes. *J. Proteome Res.* **6**, 792–800 (2007).
55. Ding, C. *et al.* Quantitative analysis of cohesin complex stoichiometry and SMC3 modification-dependent protein interactions. *J. Proteome Res.* **10**, 3652–3659 (2011).
56. Rout, M. P. & Kilmartin, J. V. Components of the yeast spindle and spindle pole body. *J. Cell Biol.* **111**, 1913–1927 (1990).
57. Strambio-de-Castillia, C., Blobel, G. & Rout, M. P. Isolation and characterization of nuclear envelopes from the yeast *Saccharomyces*. *J. Cell Biol.* **131**, 19–31 (1995).
58. Rout, M. P. & Strambio-de-Castillia, C. in *Cell Biology: A Laboratory Handbook 2* (ed. Celis, J. E.) 143–151 (Academic, 1998).
59. Cadene, M. & Chait, B. T. A robust, detergent-friendly method for mass spectrometric analysis of integral membrane proteins. *Anal. Chem.* **72**, 5655–5658 (2000).
60. Fenyo, D. *et al.* MALDI sample preparation: the ultra thin layer method. *J. Vis. Exp.* **192**, 192 (2007).
61. Field, H. I., Fenyo, D. & Beavis, R. C. RADARS, a bioinformatics solution that automates proteome mass spectral analysis, optimises protein identification, and archives data in a relational database. *Proteomics* **2**, 36–47 (2002).
62. Cox, J. & Mann, M. MaxQuant enables high peptide identification rates, individualized p.p.b.-range mass accuracies and proteome-wide protein quantification. *Nat. Biotechnol.* **26**, 1367–1372 (2008).
63. Schwanhäusser, B. *et al.* Global quantification of mammalian gene expression control. *Nature* **473**, 337–342 (2011).
64. Contino, N. C., Pierson, E. E., Keifer, D. Z. & Jarrold, M. F. Charge detection mass spectrometry with resolved charge states. *J. Am. Soc. Mass Spectrom.* **24**, 101–108 (2013).
65. Shi, Y. *et al.* A strategy for dissecting the architectures of native macromolecular assemblies. *Nat. Methods* **12**, 1135–1138 (2015).
66. Leitner, A. *et al.* Expanding the chemical cross-linking toolbox by the use of multiple proteases and enrichment by size exclusion chromatography. *Mol. Cell. Proteomics* **11**, M111.014126 (2012).
67. Yang, B. *et al.* Identification of cross-linked peptides from complex samples. *Nat. Methods* **9**, 904–906 (2012).
68. Ceyher, M. A. *et al.* Reconstitution of active human core Mediator complex reveals a critical role of the MED14 subunit. *Nat. Struct. Mol. Biol.* **21**, 1028–1034 (2014).
69. Sun, J. *et al.* The architecture of a eukaryotic replisome. *Nat. Struct. Mol. Biol.* **22**, 976–982 (2015).
70. Zheng, S. Q. *et al.* UCSF tomography: an integrated software suite for real-time electron microscopic tomographic data collection, alignment, and reconstruction. *J. Struct. Biol.* **157**, 138–147 (2007).
71. Kremer, J. R., Mastronarde, D. N. & McIntosh, J. R. Computer visualization of three-dimensional image data using IMOD. *J. Struct. Biol.* **116**, 71–76 (1996).

72. Galaz-Montoya, J. G., Flanagan, J., Schmid, M. F. & Ludtke, S. J. Single particle tomography in EMAN2. *J. Struct. Biol.* **190**, 279–290 (2015).
73. Galaz-Montoya, J. G. *et al.* Alignment algorithms and per-particle CTF correction for single particle cryo-electron tomography. *J. Struct. Biol.* **194**, 383–394 (2016).
74. Yang, Q., Rout, M. P. & Akey, C. W. Three-dimensional architecture of the isolated yeast nuclear pore complex: functional and evolutionary implications. *Mol. Cell* **1**, 223–234 (1998).
75. Beck, M. *et al.* Nuclear pore complex structure and dynamics revealed by cryoelectron tomography. *Science* **306**, 1387–1390 (2004).
76. von Appen, A. & Beck, M. Structure determination of the nuclear pore complex with three-dimensional cryo electron microscopy. *J. Mol. Biol.* **428**, 2001–2010 (2016).
77. Ludtke, S. J. Single-particle refinement and variability analysis in EMAN2.1. *Methods Enzymol.* **579**, 159–189 (2016).
78. Iwanczyk, J. *et al.* Structure of the Bim10–20 S proteasome complex by cryo-electron microscopy. Insights into the mechanism of activation of mature yeast proteasomes. *J. Mol. Biol.* **363**, 648–659 (2006).
79. Elad, N. *et al.* The dynamic conformational landscape of γ -secretase. *J. Cell Sci.* **128**, 589–598 (2015).
80. Kucukelbir, A., Sigworth, F. J. & Tagare, H. D. Quantifying the local resolution of cryo-EM density maps. *Nat. Methods* **11**, 63–65 (2014).
81. Bharat, T. A. & Scheres, S. H. Resolving macromolecular structures from electron cryo-tomography data using subtomogram averaging in RELION. *Nat. Protocols* **11**, 2054–2065 (2016).
82. Bharat, T. A., Russo, C. J., Löwe, J., Passmore, L. A. & Scheres, S. H. Advances in single-particle electron cryomicroscopy structure determination applied to sub-tomogram averaging. *Structure* **23**, 1743–1753 (2015).
83. Sampathkumar, P. *et al.* Atomic structure of the nuclear pore complex targeting domain of a Nup116 homologue from the yeast, *Candida glabrata*. *Proteins* **80**, 2110–2116 (2012). 10.1002/prot.24102
84. Sampathkumar, P. *et al.* Structure of the C-terminal domain of *Saccharomyces cerevisiae* Nup133, a component of the nuclear pore complex. *Proteins* **79**, 1672–1677 (2011).
85. Sampathkumar, P. *et al.* Structures of the autoproteolytic domain from the *Saccharomyces cerevisiae* nuclear pore complex component, Nup145. *Proteins* **78**, 1992–1998 (2010). 10.1002/prot.22707
86. Sampathkumar, P. *et al.* Structure, dynamics, evolution, and function of a major scaffold component in the nuclear pore complex. *Structure* **21**, 560–571 (2013).
87. Edelstein, A. D. *et al.* Advanced methods of microscope control using μ Manager software. *J. Biol. Methods* **1**, e10 (2014).
88. Ludtke, S. J., Baldwin, P. R. & Chiu, W. EMAN: semiautomated software for high-resolution single-particle reconstructions. *J. Struct. Biol.* **128**, 82–97 (1999).
89. Yang, Z., Fang, J., Chittiluru, J., Asturias, F. J. & Penczek, P. A. Iterative stable alignment and clustering of 2D transmission electron microscope images. *Structure* **20**, 237–247 (2012).
90. Russel, D. *et al.* Putting the pieces together: integrative modeling platform software for structure determination of macromolecular assemblies. *PLoS Biol.* **10**, e1001244 (2012).
91. Schneidman-Duhovny, D., Pellarin, R. & Sali, A. Uncertainty in integrative structural modeling. *Curr. Opin. Struct. Biol.* **28**, 96–104 (2014).
92. Sali, A. *et al.* Outcome of the first wwPDB Hybrid/Integrative Methods Task Force Workshop. *Structure* **23**, 1156–1167 (2015).
93. LoPiccolo, J. *et al.* Assembly and molecular architecture of the phosphoinositide 3-kinase p85 α homodimer. *J. Biol. Chem.* **290**, 30390–30405 (2015).
94. Luo, J. *et al.* Architecture of the human and yeast general transcription and DNA repair factor TFIIH. *Mol. Cell* **59**, 794–806 (2015).
95. Robinson, P. J. *et al.* Molecular architecture of the yeast Mediator complex. *eLife* **4**, e08719 (2015).
96. Webb, B. *et al.* Modeling of proteins and their assemblies with the Integrative Modeling Platform. *Methods Mol. Biol.* **1091**, 277–295 (2014).
97. Cherry, J. M. *et al.* *Saccharomyces* Genome Database: the genomics resource of budding yeast. *Nucleic Acids Res.* **40**, D700–D705 (2012).
98. Gautier, R., Douguet, D., Antonny, B. & Drin, G. HELIQUEST: a web server to screen sequences with specific α -helical properties. *Bioinformatics* **24**, 2101–2102 (2008).
99. Niepel, M. *et al.* The nuclear basket proteins Mlp1p and Mlp2p are part of a dynamic interactome including Esc1p and the proteasome. *Mol. Biol. Cell* **24**, 3920–3938 (2013).
100. Sali, A. & Blundell, T. L. Comparative protein modelling by satisfaction of spatial restraints. *J. Mol. Biol.* **234**, 779–815 (1993).
101. Söding, J., Biegert, A. & Lupas, A. N. The HHpred interactive server for protein homology detection and structure prediction. *Nucleic Acids Res.* **33**, W244–W248 (2005).
102. Söding, J. Protein homology detection by HMM–HMM comparison. *Bioinformatics* **21**, 951–960 (2005).
103. Buchan, D. W., Minneci, F., Nugent, T. C., Bryson, K. & Jones, D. T. Scalable web services for the PSIPRED protein analysis workbench. *Nucleic Acids Res.* **41**, W349–W357 (2013).
104. Jones, D. T. Protein secondary structure prediction based on position-specific scoring matrices. *J. Mol. Biol.* **292**, 195–202 (1999).
105. Ward, J. J., McGuffin, L. J., Bryson, K., Buxton, B. F. & Jones, D. T. The DISOPRED server for the prediction of protein disorder. *Bioinformatics* **20**, 2138–2139 (2004).
106. Marsden, R. L., McGuffin, L. J. & Jones, D. T. Rapid protein domain assignment from amino acid sequence using predicted secondary structure. *Protein Sci.* **11**, 2814–2824 (2002).
107. Lupas, A., Van Dyke, M. & Stock, J. Predicting coiled coils from protein sequences. *Science* **252**, 1162–1164 (1991).
108. Trigg, J., Gutwin, K., Keating, A. E. & Berger, B. Multicoil2: predicting coiled coils and their oligomerization states from sequence in the twilight zone. *PLoS ONE* **6**, e23519 (2011).
109. Algret, R. *et al.* Molecular architecture and function of the SEA complex, a modulator of the TORC1 pathway. *Mol. Cell. Proteomics* **13**, 2855–2870 (2014).
110. Raveh, B. *et al.* Slide-and-exchange mechanism for rapid and selective transport through the nuclear pore complex. *Proc. Natl Acad. Sci. USA* **113**, E2489–E2497 (2016).
111. Schneidman-Duhovny, D., Hammel, M. & Sali, A. FoXS: a web server for rapid computation and fitting of SAXS profiles. *Nucleic Acids Res.* **38**, W540–W544 (2010).
112. Schneidman-Duhovny, D., Kim, S. J. & Sali, A. Integrative structural modeling with small angle X-ray scattering profiles. *BMC Struct. Biol.* **12**, 17 (2012).
113. Kawabata, T. Multiple subunit fitting into a low-resolution density map of a macromolecular complex using a Gaussian mixture model. *Biophys. J.* **95**, 4643–4658 (2008).
114. Jonić, S. *et al.* Denoising of high-resolution single-particle electron-microscopy density maps by their approximation using three-dimensional Gaussian functions. *J. Struct. Biol.* **194**, 423–433 (2016).
115. Hanot, S. *et al.* Multi-scale Bayesian modeling of cryo-electron microscopy density maps. *Preprint at* <https://www.biorxiv.org/content/early/2018/02/09/113951> (2017).
116. Rout, M. P. *et al.* The yeast nuclear pore complex: composition, architecture, and transport mechanism. *J. Cell Biol.* **148**, 635–651 (2000).
117. Shen, M. Y. & Sali, A. Statistical potential for assessment and prediction of protein structures. *Protein Sci.* **15**, 2507–2524 (2006).
118. Campelo, F. & Kozlov, M. M. Sensing membrane stresses by protein insertions. *PLoS Comput. Biol.* **10**, e1003556 (2014).
119. Viswanath, S., Chennam, I. E., Cimermancic, P. & Sali, A. Assessing exhaustiveness of stochastic sampling for integrative modeling of macromolecular structures. *Biophys. J.* **113**, 2344–2353 (2017).
120. Siegel, S. *Nonparametric Statistics for the Behavioral Sciences* (McGraw-Hill, 1956).
121. McCarroll, D. *Simple Statistical Tests for Geography* (Chapman and Hall/CRC, 2016).
122. McDonald, J. H. *Handbook of Biological Statistics 3rd edn* (Sparky House, 2014).
123. Daura, X. *et al.* Peptide folding: when simulation meets experiment. *Angew. Chem. Int. Ed. Engl.* **38**, 236–240 (1999).
124. Kuzmanic, A. & Zagrovic, B. Determination of ensemble-average pairwise root mean-square deviation from experimental B-factors. *Biophys. J.* **98**, 861–871 (2010).
125. Read, R. J. *et al.* A new generation of crystallographic validation tools for the Protein Data Bank. *Structure* **19**, 1395–1412 (2011).
126. Montelione, G. T. *et al.* Recommendations of the wwPDB NMR Validation Task Force. *Structure* **21**, 1563–1570 (2013).
127. Henderson, R. *et al.* Outcome of the first Electron Microscopy Validation Task Force meeting. *Structure* **20**, 205–214 (2012).
128. Trewhella, J. *et al.* Report of the wwPDB Small-Angle Scattering Task Force: data requirements for biomolecular modeling and the PDB. *Structure* **21**, 875–881 (2013).
129. Leitner, A. *et al.* Chemical cross-linking/mass spectrometry targeting acidic residues in proteins and protein complexes. *Proc. Natl Acad. Sci. USA* **111**, 9455–9460 (2014).
130. Erzberger, J. P. *et al.* Molecular architecture of the 40S-eIF1-eIF3 translation initiation complex. *Cell* **158**, 1123–1135 (2014).
131. Alber, F., Förster, F., Korkin, D., Topf, M. & Sali, A. Integrating diverse data for structure determination of macromolecular assemblies. *Annu. Rev. Biochem.* **77**, 443–477 (2008).
132. Alber, F., Chait, B. T., Rout, M. P. & Sali, A. in *Protein–Protein Interactions and Networks: Identification, Characterization and Prediction* (eds Panchenko, A. & Przytycka, T.) 99–114 (Springer, 2008).
133. Ermak, D. L. & McCammon, J. A. Brownian dynamics with hydrodynamic interactions. *J. Chem. Phys.* **69**, 1352–1360 (1978).
134. Hough, L. E. *et al.* The molecular mechanism of nuclear transport revealed by atomic-scale measurements. *eLife* **4**, e10027 (2015).
135. Milles, S. *et al.* Plasticity of an ultrafast interaction between nucleoporins and nuclear transport receptors. *Cell* **163**, 734–745 (2015).
136. Sakiyama, Y., Mazur, A., Kapinos, L. E. & Lim, R. Y. Spatiotemporal dynamics of the nuclear pore complex transport barrier resolved by high-speed atomic force microscopy. *Nat. Nanotechnol.* **11**, 719–723 (2016).
137. van der Maarel, J. R. C. *Introduction to Biopolymer Physics* (World Scientific, 2008).
138. Denning, D. P., Patel, S. S., Uversky, V., Fink, A. L. & Rexach, M. Disorder in the nuclear pore complex: the FG repeat regions of nucleoporins are natively unfolded. *Proc. Natl Acad. Sci. USA* **100**, 2450–2455 (2003).
139. Lemke, E. A. The multiple faces of disordered nucleoporins. *J. Mol. Biol.* **428**, 2011–2024 (2016).
140. Lim, R. Y. *et al.* Flexible phenylalanine–glycine nucleoporins as entropic barriers to nucleocytoplasmic transport. *Proc. Natl Acad. Sci. USA* **103**, 9512–9517 (2006).

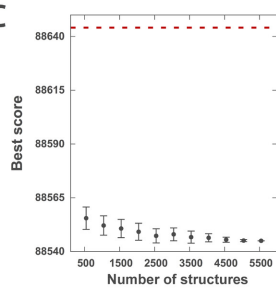
A



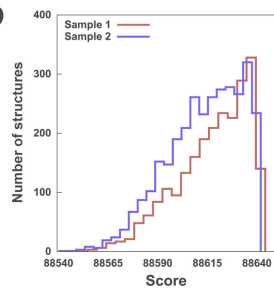
B



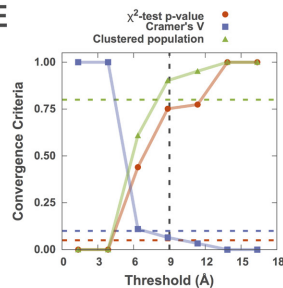
C



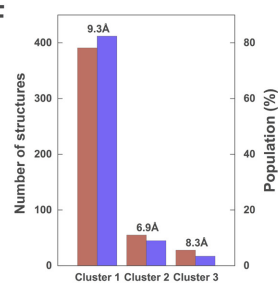
D



E



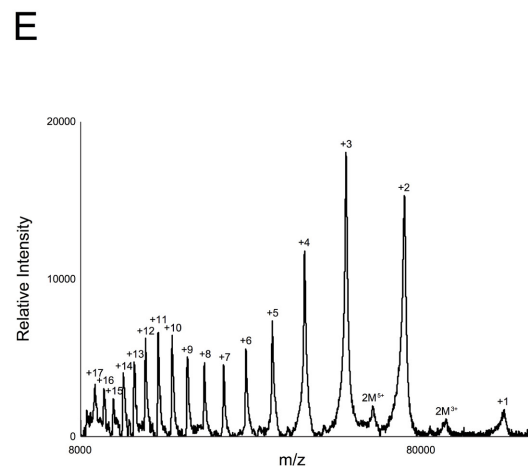
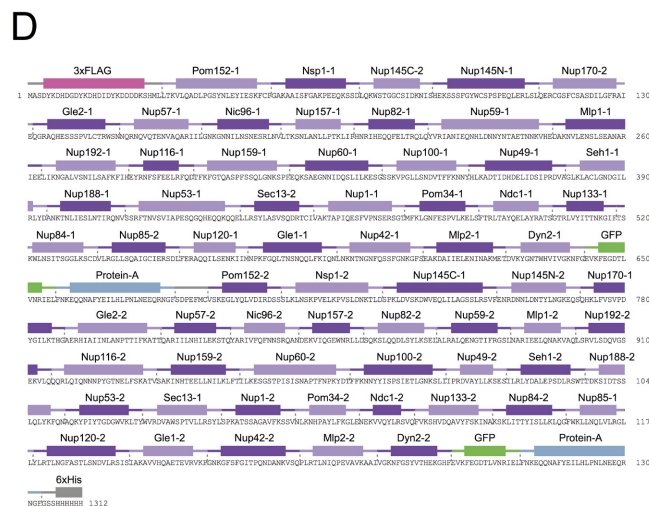
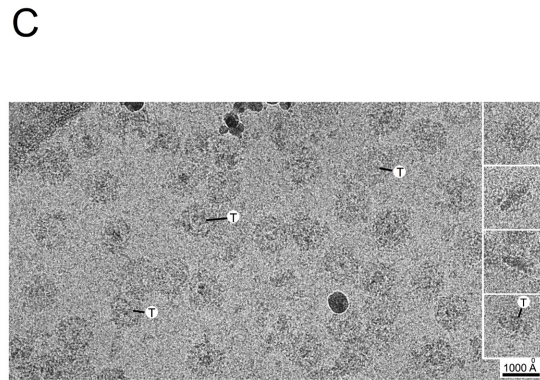
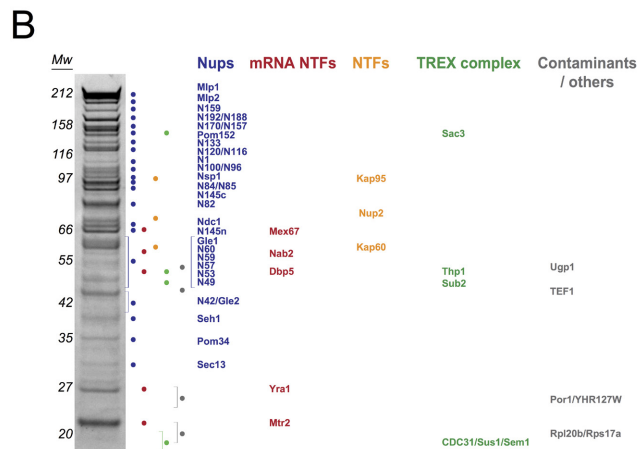
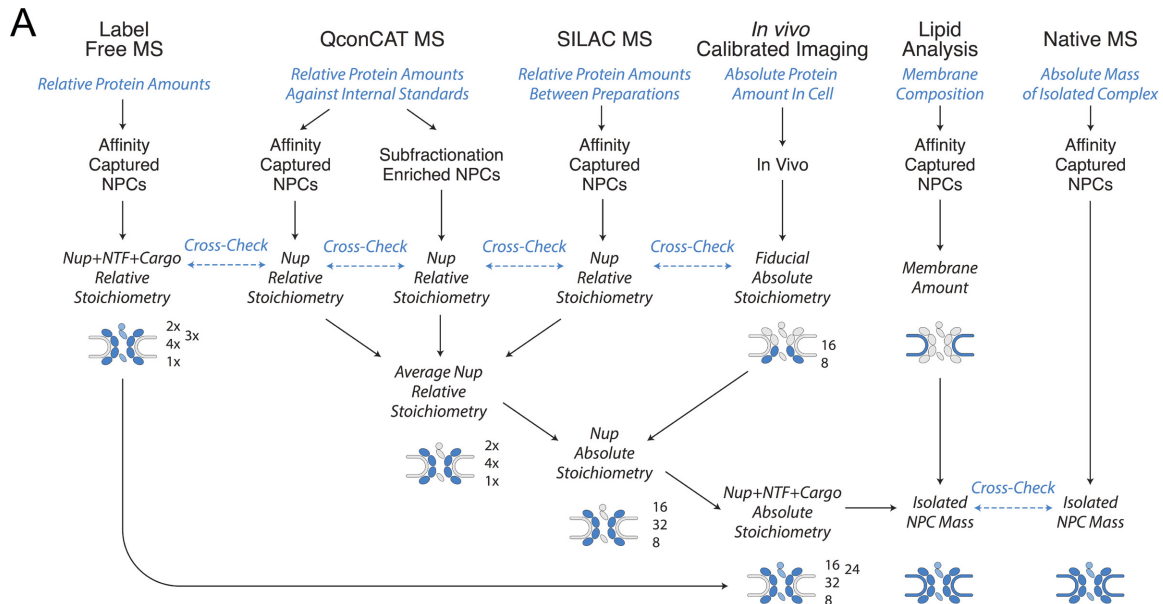
F



Extended Data Figure 1 | See next page for caption.

Extended Data Figure 1 | Integrative structure determination of the *S. cerevisiae* NPC at 9 Å precision. **a**, Schematic of integrative structure determination of the *S. cerevisiae* NPC. Random initial structures of the Nups and their sub-complexes were optimized by satisfying spatial restraints implied by the input information. **b**, The full description of integrative structure determination of the *S. cerevisiae* NPC, proceeded through four stages^{8,90–92} (Supplementary Table 3): (1) gathering data, (2) representing subunits and translating data into spatial restraints, (3) configurational sampling to produce an ensemble of structures that satisfies the restraints and (4) analysing and validating the ensemble structures and data (Extended Data Figs 7, 8 Supplementary Tables 2–4 and Methods). The integrative structure modelling protocol (stages 2, 3 and 4) was scripted using the Python modelling interface package version 4d97507, which is a library for modelling macromolecular complexes based on our open-source IMP package⁹⁰ version 2.6 (<https://integrativemodeling.org>). **c**, Convergence of the structure score for the 5,529 good-scoring NPC structures; the scores do not continue to improve as more structures are computed, essentially independently of each other. The error bar represents the standard deviations of the best scores, estimated by repeating sampling of NPC structures ten times ($n = 10$, mean score values plotted). The red dotted line indicates the total score threshold (88,644.1) that defines the good-scoring NPC structures (Methods). **d**, Distribution of scores for structure samples 1 (red) and 2 (blue), comprising the 5,529 good-scoring NPC structures ($n^{\text{sample1}} = 2,359$

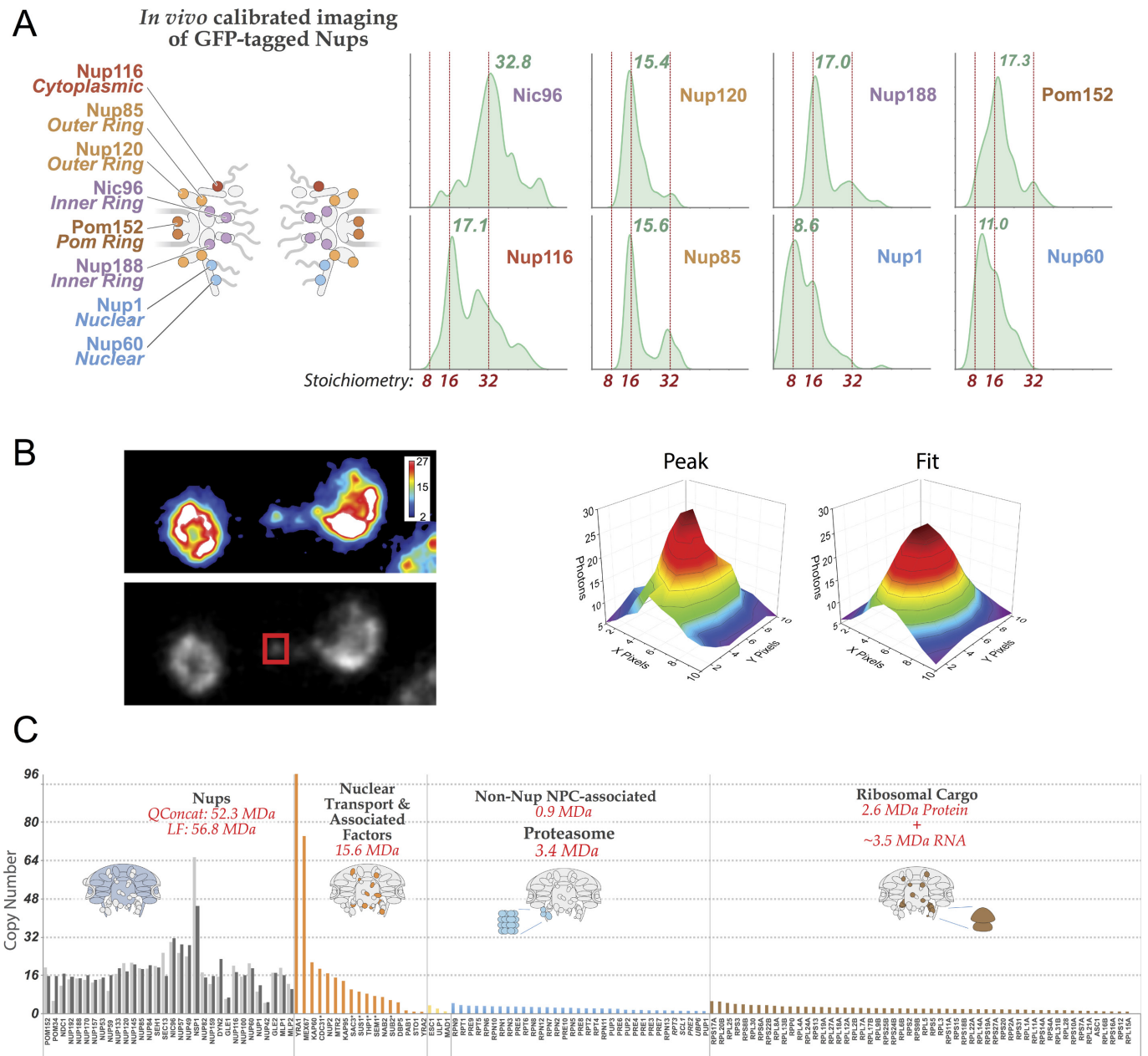
and $n^{\text{sample2}} = 3,170$ structures). The non-parametric Kolmogorov–Smirnov two-sample test^{120,121} (two-sided) indicates that the difference between the two score distributions is insignificant (P value (1.0) > 0.05). In addition, the magnitude of the difference is small, as demonstrated by the Kolmogorov–Smirnov two-sample test statistic, D , of 0.045. Thus, the two score distributions are effectively equal. **e**, Three criteria for determining the sampling precision (y axis), evaluated as a function of the r.m.s.d. clustering threshold¹²³ (x axis) ($n = 5,529$ structures). First, the P value is computed using the χ^2 -test (one-sided) for homogeneity of proportions¹²² (red dots). Second, an effect size for the χ^2 -test is quantified by the Cramer's V value (blue squares). Third, the population of structures in sufficiently large clusters (containing at least ten structures from each sample) is shown as green triangles. The vertical dotted grey line indicates the r.m.s.d. clustering threshold at which three conditions are satisfied (χ^2 -test P value (0.75) > 0.05 (red, horizontal dotted line), Cramer's V (0.065) < 0.10 (blue, horizontal dotted line) and the population of clustered structures (0.90) > 0.80 (green, horizontal dotted line)), thus defining the sampling precision of 9 Å. The three solid curves (in red, blue and green) were drawn through the points to help visualize the results. **f**, Population of sample 1 and 2 structures in the three clusters obtained by threshold-based clustering¹²³ using an r.m.s.d. threshold of 12 Å. The dominant cluster (cluster 1) contains 80.3% of the structures. Cluster precision is shown for each cluster. The precision of the dominant cluster defines the structure precision.



Extended Data Figure 2 | See next page for caption.

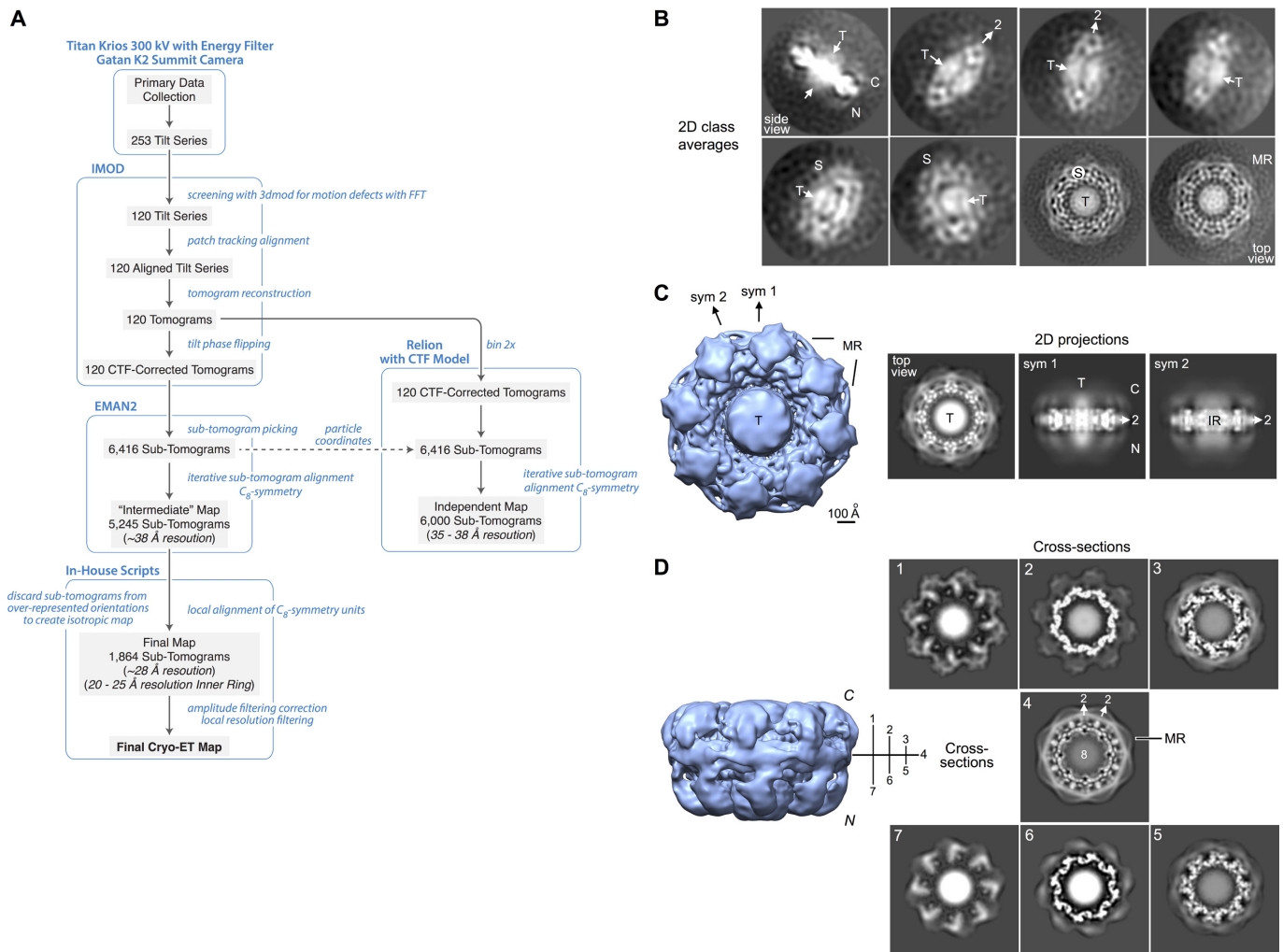
Extended Data Figure 2 | Quantitative analysis of the mass and stoichiometry of the endogenous NPC (part 1). **a**, A multipronged approach to accurately define the mass, stoichiometry and composition of native macromolecular assemblies. Schematic is shown of the multiple orthologous methods that are integrated within our strategy for the analysis of native assemblies. The main experimental methods are listed on top, followed by the characteristic that they help to quantify (in blue) and the type of sample to which they were applied. The final outcome of each method is indicated (black arrows); the steps of each method that are compared to serve as a cross-check control are indicated (blue dashed lines). At bottom, the integration of the different data points into a final comprehensive description of the endogenous assembly is depicted. Small cartoon insets of the NPC are used to illustrate the analysis. **b**, SDS-PAGE analysis of the affinity-captured *S. cerevisiae* NPCs isolated from an MLP1-PPX-PrA tagged strain ($n > 20$ independent experiments). Molecular weight marker values (Mw) are indicated to the left of the gel lane. Dots signify the main protein components of the isolated NPCs as identified by liquid chromatography-mass spectrometry (Extended Data Fig. 3c). Proteins are grouped and coloured by functional categories or membership of discrete macromolecular assemblies. Nups, blue; mRNA transport factors (TFs), red; transport factors, orange; transcription-

export (TREX), green; contaminants and/or others, grey. For gel source data, see Supplementary Fig. 1. **c**, Cryo-electron microscopy analysis of the affinity-captured NPCs ($n > 20$ independent experiments). The particles have a clear preferred orientation (Methods). Some side views are presented in the inset. The central transporter is present in every NPC (indicated by 'T'). Scale bar, 1,000 Å. **d**, Schematic showing the primary amino acid sequence of the 148.2 kDa synthetic QconCAT-A. It includes two peptides for each Nup (thick bars), arranged in the indicated order. The native three amino acid residues flanking regions (thin bars) of each peptide were included to preserve the native trypsin target sequence. A N-terminal 3×FLAG tag was included, as well as a C-terminal 6×His tag for purification under denaturing conditions. The stringent criteria used for the selection of the QconCAT peptides are described in the Methods. **e**, MALDI mass spectrometry spectrum of intact, purified full-length QconCAT-A labelled with stable isotope, showing that a single species was detected. Numbers above peaks denote the QconCAT-A protein species with n positive charges and the 2M QconCAT-A protein dimer. The measured molecular weight of the QconCAT-A labelled with stable isotope was $149,049 \pm 38$ Da, consistent with its calculated molecular weight of 148,200 Da (Methods).



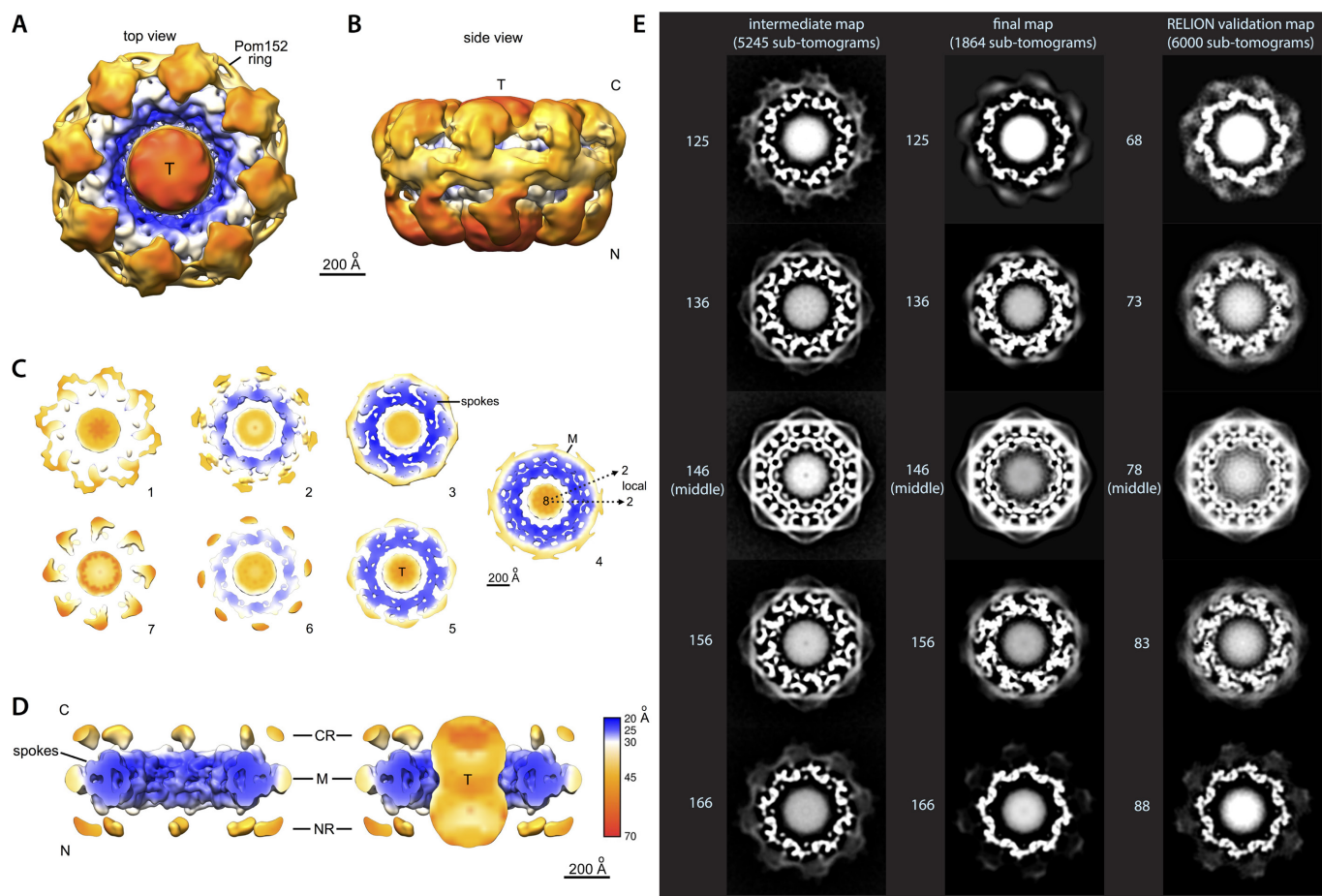
Extended Data Figure 3 | Quantitative analysis of the mass and stoichiometry of the endogenous NPC (part 2). **a**, Left, schematic localization of the Nup-GFP reporters selected for the *in vivo* calibrated imaging stoichiometry analyses. Nups were selected to represent every major NPC module and to provide comprehensive coverage of the assembly. Right, Kernel density estimation of distributions of GFP proteins per Nup were calculated from the calibrated imaging data. $n = 48-178$. **b**, Heat map of a yeast cell expressing Nup85-GFP. Image (left) was acquired as described in Methods. In addition, for illustration purposes, a maximum projection along the z axis was performed, and the image was smoothed with a Gaussian blur of radius 1. A heat map

was used to illustrate intensity units in raw photon counts. For the area outlined in a red rectangle, a 2D distribution of photon counts and the corresponding Gaussian fit are shown (right). **c**, Stoichiometries of main components associated with the affinity-captured NPCs, as determined by label-free mass spectrometry quantification (at least three peptides per protein). Proteins are grouped and coloured by functional categories or membership of discrete macromolecular assemblies. The TREX complex components are included in the 'Transport & Associated Factors' category and labelled with an asterisk. QconCAT-derived stoichiometries for all the Nups (dark grey bars) are shown for comparison.



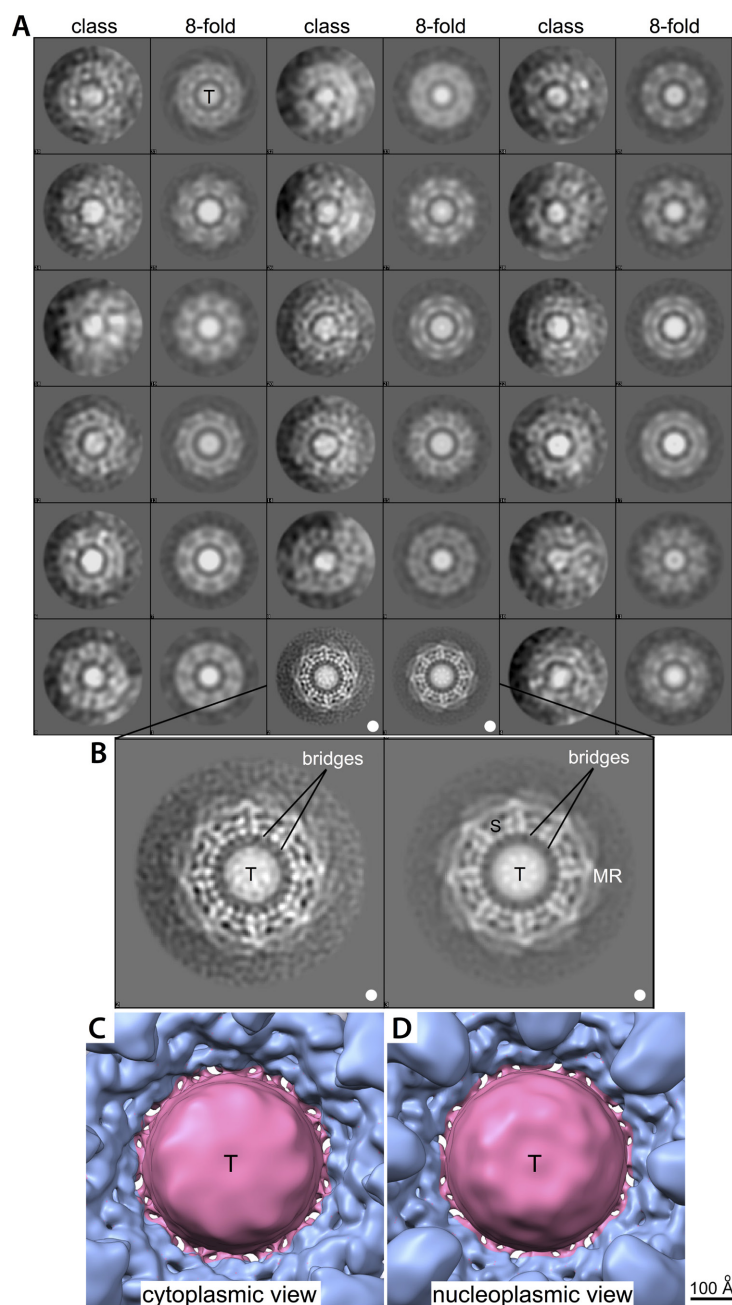
Extended Data Figure 4 | Cryo-ET strategy and the resulting 3D cryo-ET map of the NPC exhibiting non-enforced local C₂-symmetry axes in the inner ring. **a**, Diagram describing the methodology used to obtain the whole *S. cerevisiae* NPC cryo-ET map (Methods). **b**, 2D class averages are shown (protein in white), which were calculated using the original unaligned sub-tomograms projected along the *z* axis. The overall thickness of the *S. cerevisiae* NPC is apparent in a side-view class, and the local C₂-symmetry axes in the inner ring are also apparent (indicated with '2'). The transporter density is present in every class. **c**, Left, top view of the cryo-ET map with the two local C₂-symmetry axes indicated by

arrows and labels (sym 1 and sym 2). They are 22.5° apart, owing to the C₈-symmetry axis. Right, 2D projections of the top view, and two side views along the two local C₂-symmetry axes (side 1 and side 2 projected along axes sym 1 and sym 2, respectively). **d**, Seven cross-sections of the cryo-ET map are shown on the right (labelled 1–7) with their positions in the 3D map indicated in the side view on the left. The local C₂-symmetry of the inner ring is apparent in cross-sections 2–6, mirrored about the central section in panel 4. Labels throughout: C, cytoplasm; N, nucleoplasm; T, central transporter; S, core scaffold; MR, membrane ring; IR, inner ring. Scale bar, 100 Å.



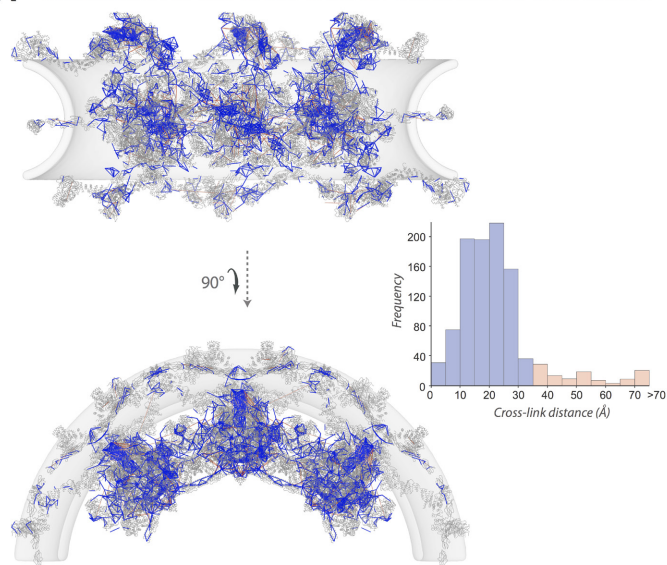
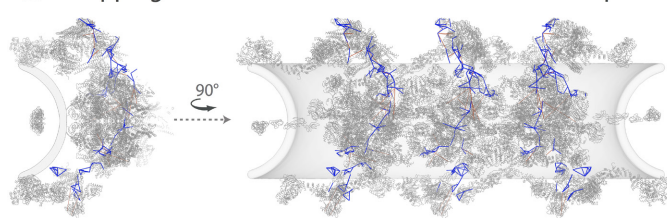
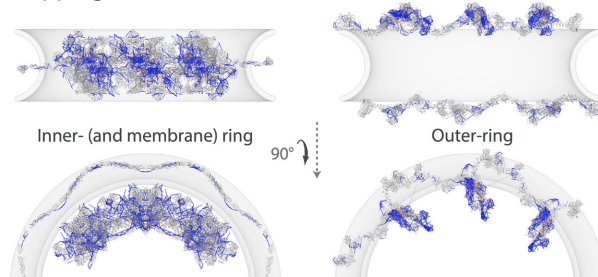
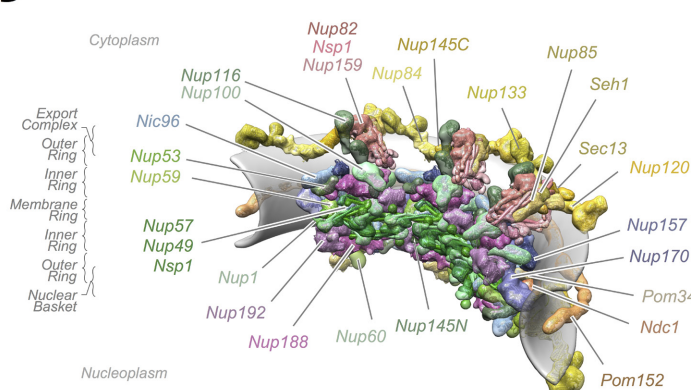
Extended Data Figure 5 | Resolution estimates for the cryo-ET map of the NPC and comparison of cross-sections between the intermediate, final and RELION cryo-ET maps. **a, b**, Top (**a**) and side (**b**) views of the cryo-ET map are colour-coded according to local resolution estimates (colour bar), and are shown at a low threshold to reveal weaker density features at the periphery of the NPC that are more flexible. **c**, Cross-sections are shown at a reduced scale, colour coded according to local resolution estimates (colour bar). A remnant of the pore membrane (M) is present, encircling the entire mid-line of the NPC. Sections 3–5 are shown at a higher threshold. In section 3, the inner-ring region is indicated by ‘spokes’. In section 4, local C_2 -symmetry axes are indicated by dashed

arrows. **d**, Thick sections of the inner ring are shown at a higher threshold, as viewed along the membrane plane. Note that the inner ring (indicated) is almost entirely in the 20–25 Å resolution range. C, cytoplasmic side; N, nuclear side; CR, cytoplasmic ring; NR, nuclear ring; M, pore membrane; T, central transporter. **e**, Comparison of five cross-sections (cross-section number on left) in the inner-ring region of the NPC between cryo-ET maps in different stages of the reconstruction process (Extended Data Fig. 4a): intermediate map (left column), final map (middle column) and an independent validation map, reconstructed with RELION at a twice-reduced Å per pixel size (right column). Details on the reconstruction of maps are provided in Methods.

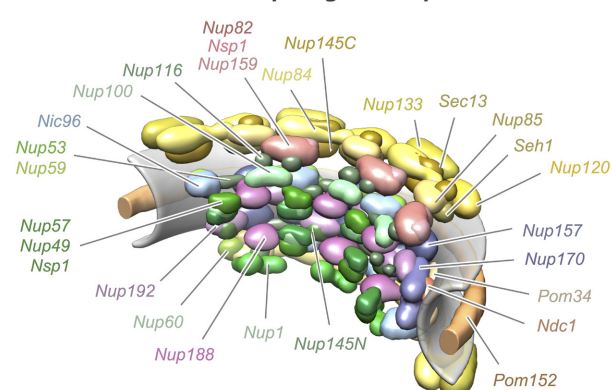
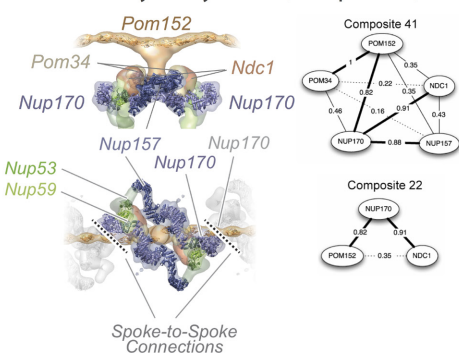
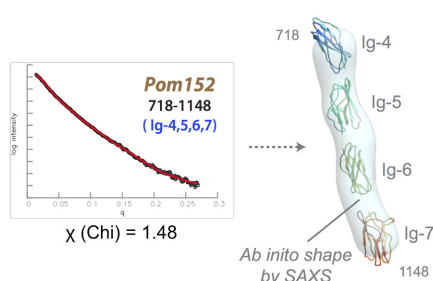
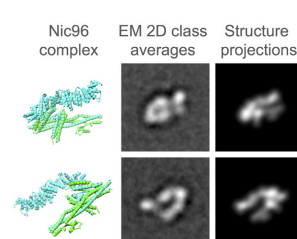


Extended Data Figure 6 | 2D classification of projections from 1,864 original NPC sub-tomograms aligned with their C_8 -symmetry axis nearly along the z axis. **a, In total, 18 good class averages are shown after maximum likelihood classification (using RELION 1.4^{81,82}) without symmetry imposed. Each class average (on the left) is paired with a C_8 -symmetry enforced image of itself (on the right). Central transporter densities are present in each of the class averages (both with and without imposed C_8 -symmetry), indicating that the central transporter is generally present in these particles. **b**, An expanded view of a large class from**

a shows bridges (indicated) between the core scaffold and the central transporter, both before (left) and after averaging (right) using the C_8 -symmetry of the NPC. S, core scaffold; MR, membrane ring; T, central transporter. Matching panels in **a** and **b** are marked with white dots. **c**, **d**, The cryo-ET 3D map is presented as in Fig. 3 and is zoomed in to show the meshwork of bridges between the scaffold and the central transporter, as viewed from the cytoplasm and the nucleoplasm, respectively.

A Satisfaction of the Cross-links on the entire NPC Structure**B** Mapping of the Cross-links onto the Connector Nups**C** Mapping of the Cross-links onto the inner- and outer-rings**D** 2017 Structure

2007 Topological Map

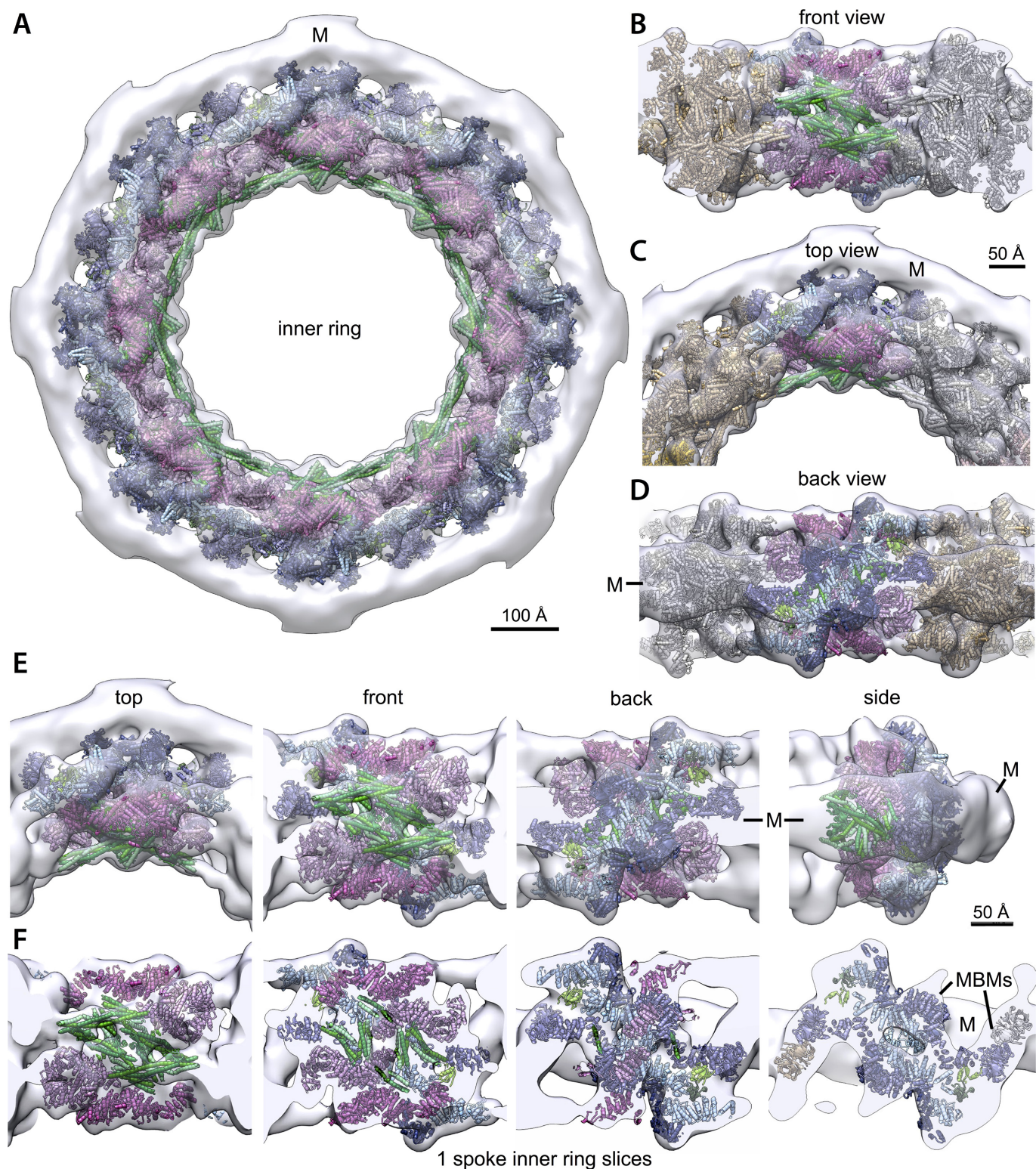
**E** Satisfaction of affinity purification and overlay assays data (composites)**F** Satisfaction of SAXS data**G** Satisfaction of EM 2D

Extended Data Figure 7 | See next page for caption.

Extended Data Figure 7 | Validation of the NPC structure (part 1).

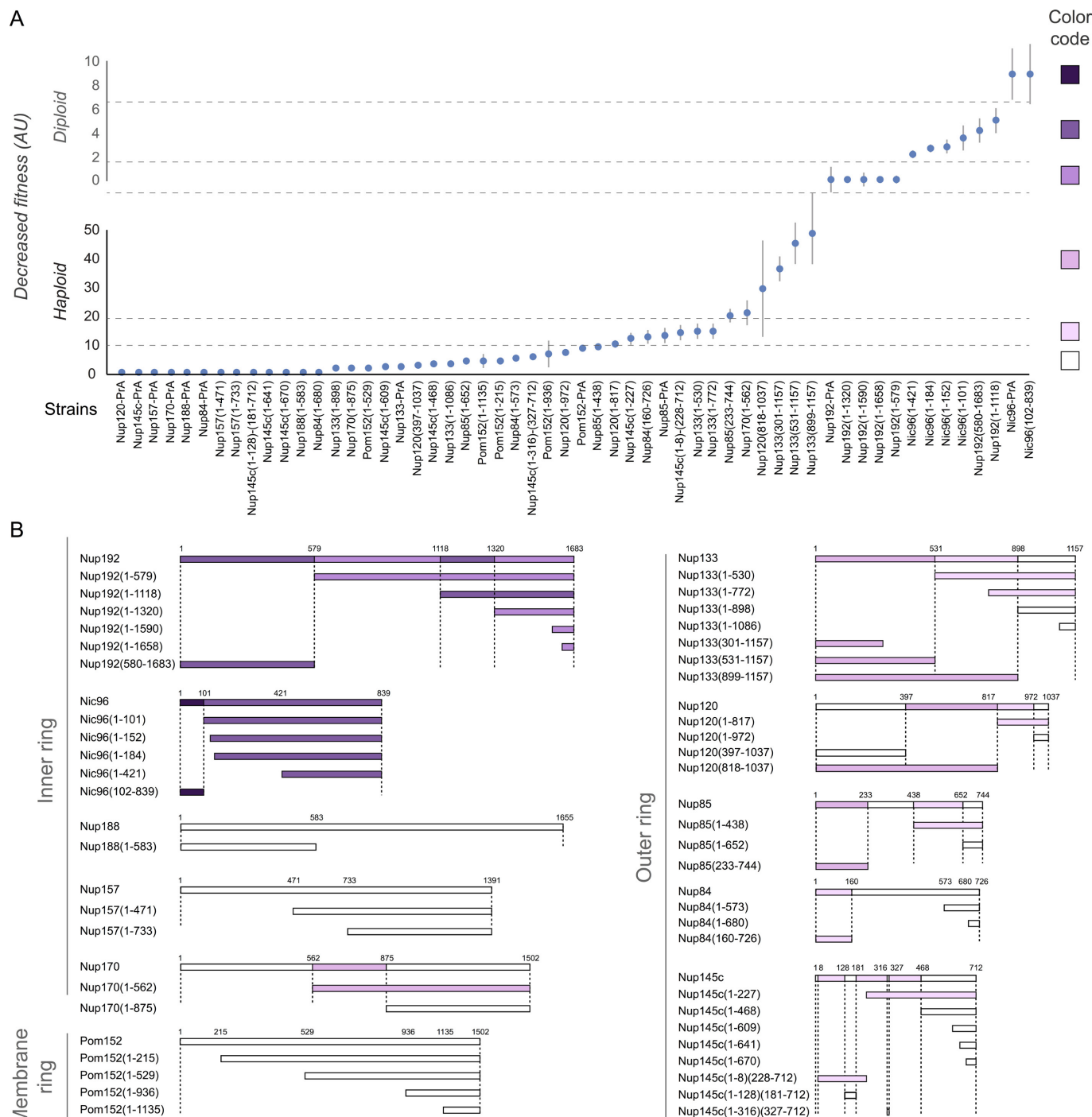
a–c, Satisfaction of the chemical cross-links. **a**, Identified chemical cross-links were mapped onto the integrative structure of the entire NPC, as shown in front (upper) and top (lower) views. Satisfied cross-links, with C_{α} – C_{α} distances that fall within the distance threshold of 35 Å in at least one good-scoring NPC structure, are shown in blue. Violated cross-links, with C_{α} – C_{α} distances that are larger than 35 Å, are shown in orange. The histogram on the right shows the distribution of the cross-linked C_{α} – C_{α} distances, validating the NPC structure. **b**, Mapping of the cross-links onto the cytoplasmic and nucleoplasmic connector Nups (Nup116, Nup100, Nup145N, Nup1 and Nup60). Front (right) and side (left) views show how the NPC outer rings are connected to the inner ring through a network of connector Nups across the length of the spoke. **c**, Mapping of the cross-links onto the inner (and membrane) and outer rings, in front (upper) and top (lower) views. **d–g**, Satisfaction of data and considerations that were not used to compute the structure. **d**, Our integrative structure of the NPC (left) was compared a previously published topological map^{3,8} (right). The two structures are consistent with each other, though our integrative structure is defined at an order of magnitude higher precision. **e**, Satisfaction of affinity purification and overlay assays data (composites); our current structure satisfies all 82 composites determined by affinity

purification and overlay assays^{3,8}, even though these data were not used in its determination. For example, Pom152, Pom34, Ndc1, Nup157 and Nup170 are connected with each other (left), consistent with the previously published composites determined using the affinity purification data^{3,8} (right). **f**, Satisfaction of SAXS data; the atomic structures of eight Nups are consistent with the corresponding SAXS profiles for their constructs^{9,12,23,83–86} (Supplementary Tables 2, 6 and Methods). For example, the SAXS profile calculated from the atomic structure of Pom152^{718–1148} (red curve) using FoXS¹¹¹ is well-matched ($\chi = 1.48$) to the corresponding experimental SAXS profile²³ (black dots; $n = 20$ exposures). For visualization purposes, the Pom152^{718–1148} structure (represented as a ribbon) is shown along with the best fit of the *ab initio* shape (represented as a transparent envelope) computed from the experimental SAXS profile. **g**, Satisfaction of the negative-stain electron microscopy 2D class averages for the Nic96 complex; the structures of the Nic96 complex (composed of Nic96, Nsp1, Nup49 and Nup57) in the dominant cluster can be projected well on 2D class averages obtained for the natively isolated complex ($n = 5,458$ particles; Methods). The experimental class averages were satisfied by the structure with cross-correlation coefficients of 0.85 and 0.80, respectively (Methods).



Extended Data Figure 8 | Validation of the NPC structure (part 2), showing consistency between the NPC structure and the cryo-ET density map. The cryo-ET density map is shown at a high-density threshold (grey) to reveal details of the inner ring. A representative structure of the inner ring is shown docked into the density, showing the excellent fit. All Nups are coloured as in Fig. 4. The pore membrane is

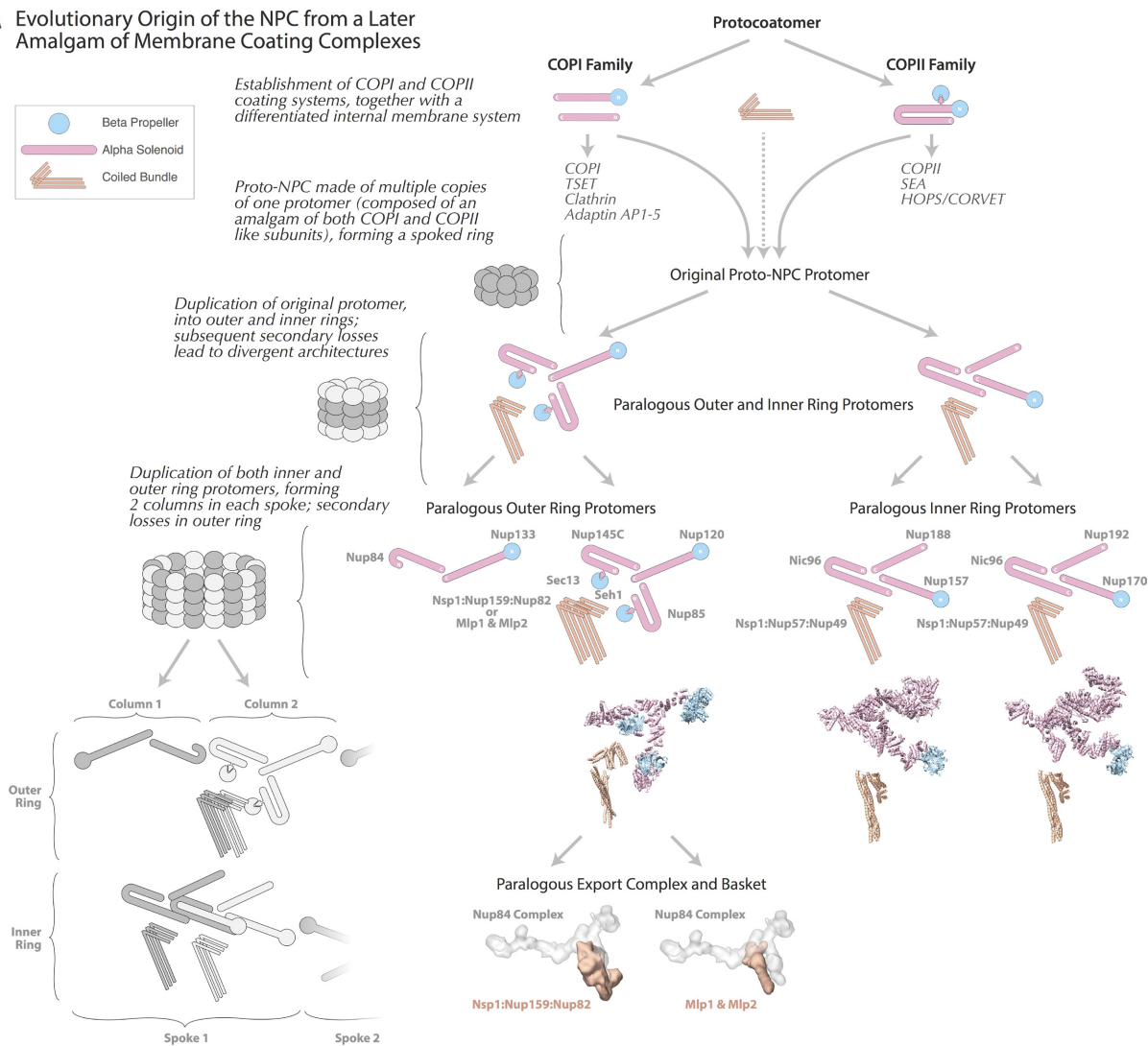
indicated by M. **a**, Full eight-spoke inner ring (scale bar, 100 Å). **b–d**, front (**b**), top (**c**) and back (**d**) views of three spokes with neighbours coloured brown and grey (scale bar, 50 Å). **e**, Different views of a single spoke (scale bar, 50 Å) are shown within the density map. **f**, Thick cross-sections are shown through a single spoke in the inner ring, as viewed from the central C_8 -symmetry axis (scale bar, 50 Å); MBMs (see Fig. 5) are indicated.



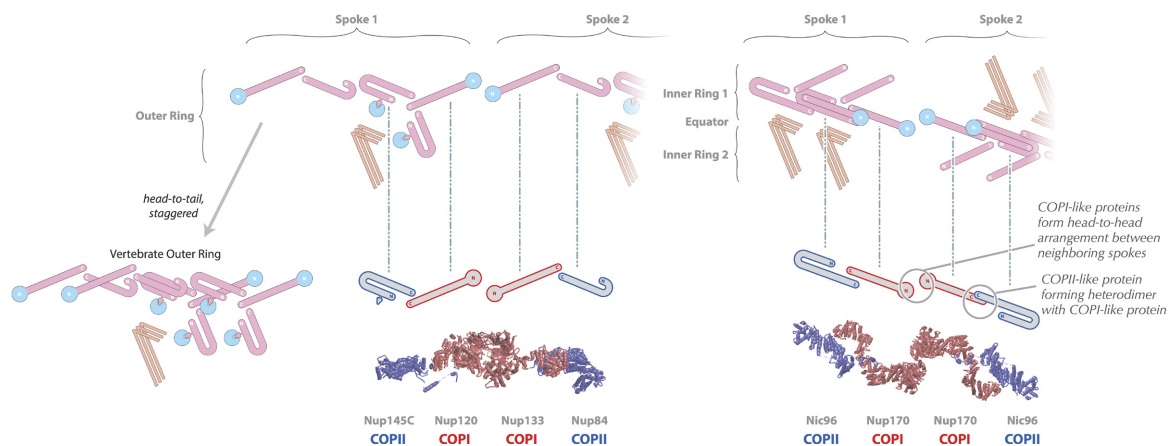
Extended Data Figure 9 | Functional analysis of the fitness of nucleoporin mutants using ODELAY. **a**, The fitness defect phenotype was quantified and plotted (mean Z-score; $n = 6$ experiments, containing at least 200–300 individuals per point; see Methods for details) for each nucleoporin truncation or C-terminal protein-A tagged mutant in order of decreasing fitness (increasing number of units), as observed by ODELAY assay¹⁴ (Methods). Strains for which truncations in a haploid background were found to lead to lethality after tetrad dissection (Nic96 and Nup192) were assigned the maximum level of defect and plotted on top of the rest

(diploid), on the basis of the fitness phenotype observed for the indicated Nic96 and Nup192 mutants in a diploid background (in which a wild-type copy of the nucleoporin is also present and expressed). Six divisions were assigned based on decreasing levels of fitness^{9,13}; white (wild type) to dark purple (severe defect). AU, arbitrary unit; error bar = standard deviation. **b**, Mapping of the colour code described in **a** into the NPC components. Horizontal lines represent the amino acid residue length of each protein and truncated version; amino acid residue positions are shown on top of the lines.

A Evolutionary Origin of the NPC from a Later Amalgam of Membrane Coating Complexes



B Conserved Structural Motifs between Outer and Inner Rings

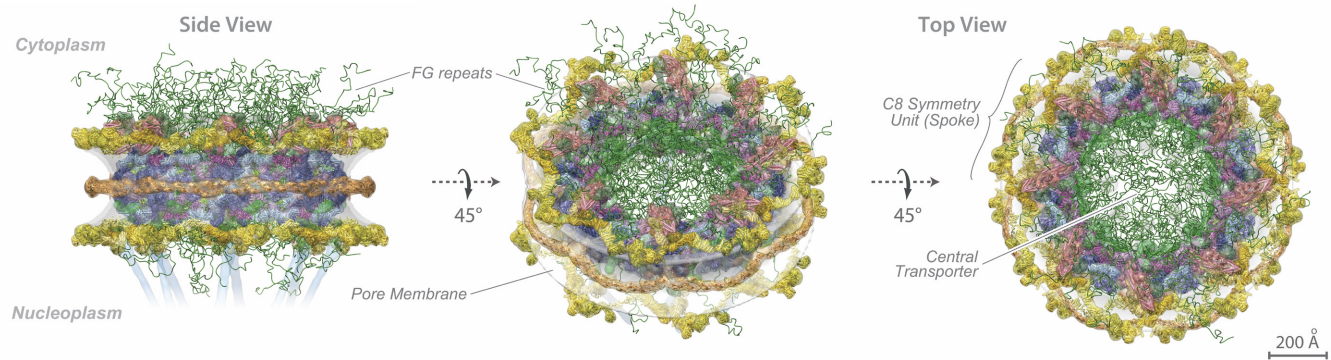


Extended Data Figure 10 | See next page for caption.

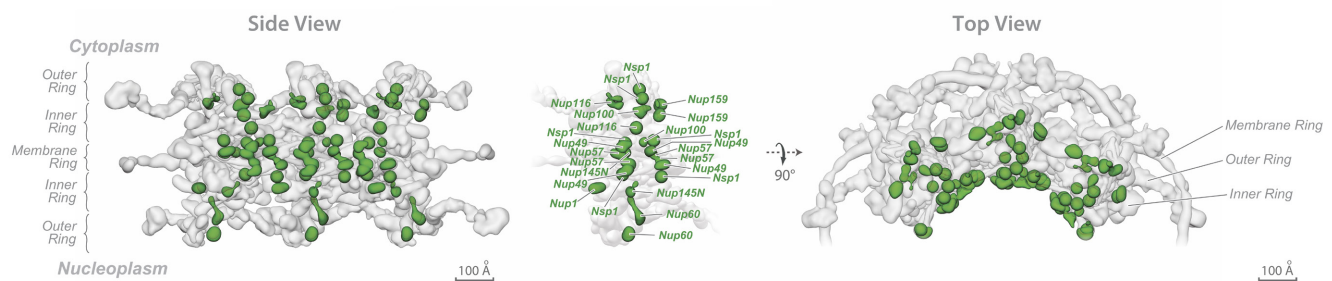
Extended Data Figure 10 | Proposed evolutionary origin of the NPC from a later amalgam of membrane coating complexes. **a**, Diagram depicting how the NPC may have originated from an ancestral coatomer module through a series of duplications, divergence and secondary loss events. Top, the origin of an ancestral proto-NPC coatomer module from an amalgamation of COPI-like and COPII-like complexes. Middle, the initial duplication leading to the origin of the inner and outer rings, and their associated coiled bundles. Presumed secondary losses removed the additional COPII-like subunit of the inner-ring protomer; loss of the adaptin-like subunit from the outer ring may have occurred here, or later in only certain lineages. Bottom, another duplication and divergence within each spoke may then have generated two parallel and laterally-offset paralogous columns; in the outer ring, a COPII-like subunit was then lost from one of the duplicates. The coiled bundles of the outer rings gave rise to the cytoplasmic export complex and nuclear basket

by subsequent duplication; the export complex itself is a duplicate with a dimer of trimeric coiled bundles in its core. Outer-ring duplications are not shown. Relevant nucleoporin domains are depicted as follows: β -propellers (cyan circles), α -solenoids (pink bars) and coiled-coil domains (orange sticks). Left, diagrams (grey) exemplify the path of duplications within the whole NPC. Examples of ribbon representations for each module are presented. The anchoring points of the coiled-coil cytoplasmic Nup82 complex and the nuclear basket (orange densities) into an equivalent region of the outer-ring Nup84 complex (grey density) are shown. **b**, Conserved structural motifs connecting spokes in the outer and inner rings. Diagram showing how the spoke-to-spoke connection is established through similar head-to-head connections of heterodimers containing one COPI-like and one COPII-like subunits in both the outer (left) and the inner (right) NPC rings. Top, nucleoporin domains coloured as in **a**; bottom, COPI-like Nups in red, COPII-like Nups in blue.

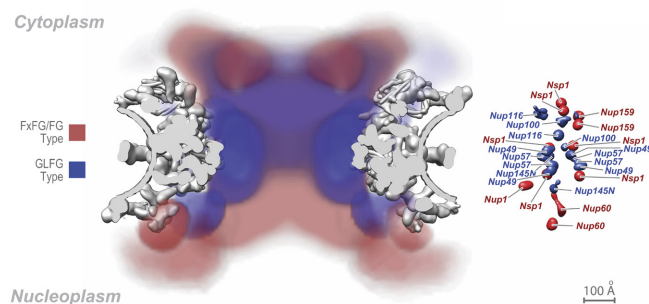
A A snapshot of FG Repeats in the complete structure of the NPC



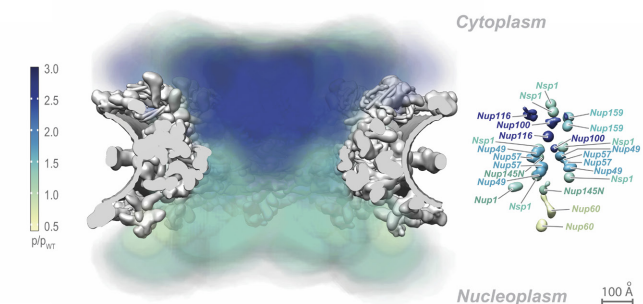
B Position of the FG Repeat Anchor Sites in the NPC



C Map of FG Repeat Types



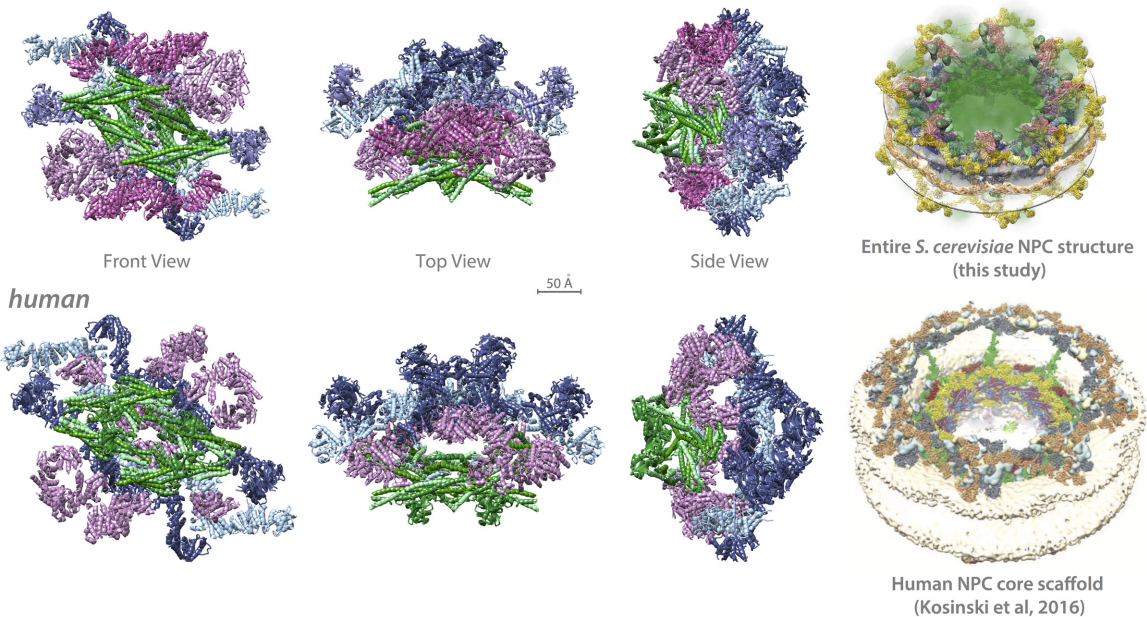
D Map of FG Repeat Contribution to Pore Permeability



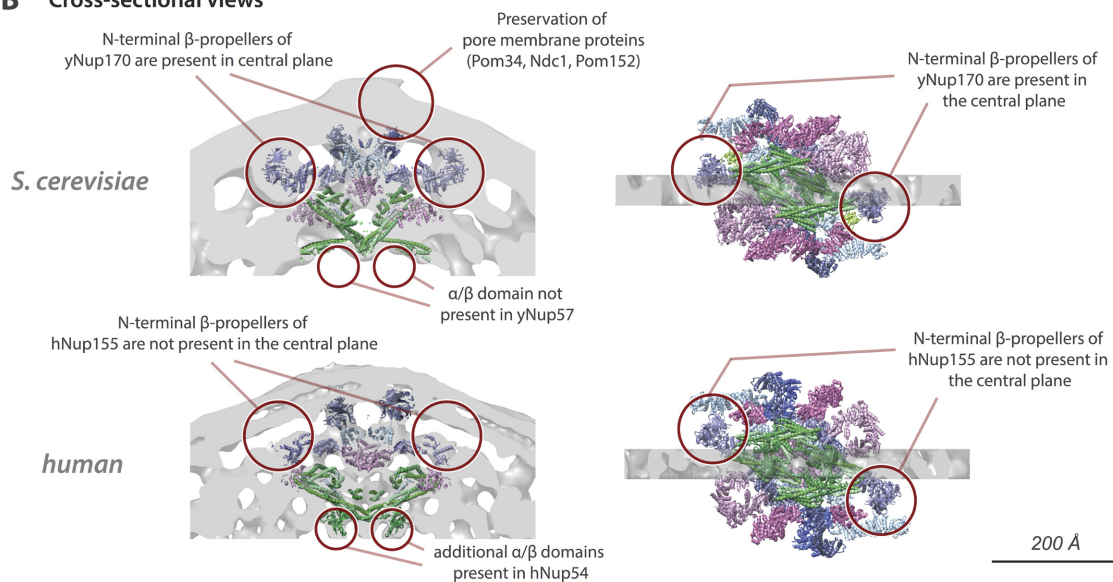
Extended Data Figure 11 | Position of the FG-repeat anchor points and heat mapping of the FG repeats. **a**, Three views of the complete structure of the NPC are shown with major structural features (coloured as in Fig. 4 and Supplementary Table 2) and a snapshot of modelled FG-repeat regions (indicated in green). For each Nup, the localization probability density of the ensemble of structures is shown with a representative structure from the ensemble embedded within it. See also Supplementary Videos 1–3. Scale bar, 200 Å. **b**, Positions of FG-repeat anchor points within the ensemble of solutions are depicted as green surfaces; the Nups to which they belong are labelled in the centre image. Left, side view of three spokes; centre, side view of one spoke; right, top view of three spokes. Scale bar,

100 Å. **c**, Heat mapping of the type of FG-repeat region of each FG Nup (FXFG/FG type, red; GLFG type, blue), showing partitioning of the FG types to different regions of the central transporter. Identity of mapped Nups is shown in the diagram on the right. Scale bar, 100 Å. **d**, Heat mapping of the effect on NPC permeability of the truncation of an FG repeat in each FG Nup, relative to the wild-type strain (p/p_{WT}); the severity of the permeability defect is indicated in increasing intensity of shades of blue from minor defect (light green) to severe defect (dark blue), thereby defining the FG repeats that are most important in maintaining the passive permeability barrier. Identities of mapped Nups are shown in the diagram on the right. Scale bar, 100 Å.

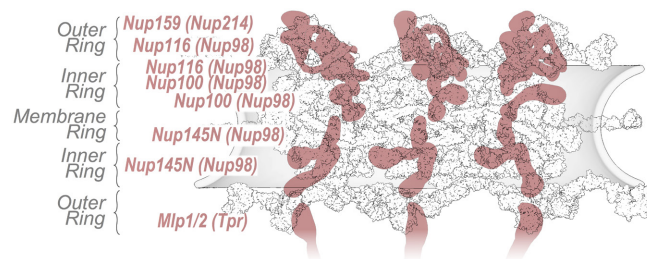
A Comparison between the NPC inner-rings in the *S. cerevisiae* structure and human core scaffold



B Cross-sectional views



C Homologs of Oncogenic Hotspots



Extended Data Figure 12 | Comparison of the *S. cerevisiae* NPC structure and human NPC core scaffold. **a**, Comparison between the inner rings in the structure of the *S. cerevisiae* NPC (first row) and the core scaffold of the human NPC (second row) (Protein Data Bank code: 5IJN⁶). Yeast Nups are coloured as in Fig. 4; human Nup homologues are coloured as their yeast counterparts. All copies of human Nup155 are coloured as yeast Nup157, and all copies of human Nup205/Nup188 are coloured as yeast Nup192. Only homologue Nups present in both yeast and human

are shown. The human NPC core scaffold includes two additional copies of Nup155 that are absent in yeast (owing to the different stoichiometry between organisms). Yeast Nup53 and Nup59 are not shown because their counterparts are not present in the human NPC core scaffold. **b**, Major differences in the inner ring between the *S. cerevisiae* and human NPCs are highlighted, in the cross-sectional view near the equator. **c**, Positions of yeast Nups homologous to oncogenic human Nups (in parentheses) are shown in red, mapped onto three spokes of the NPC.

A single population of red globular clusters around the massive compact galaxy NGC 1277

Michael A. Beasley^{1,2}, Ignacio Trujillo^{1,2}, Ryan Leaman³ & Mireia Montes⁴

Massive galaxies are thought to form in two phases: an initial collapse of gas and giant burst of central star formation, followed by the later accretion of material that builds up their stellar and dark-matter haloes^{1–4}. The systems of globular clusters within such galaxies are believed to form in a similar manner. The initial central burst forms metal-rich (spectrally red) clusters, whereas more metal-poor (spectrally blue) clusters are brought in by the later accretion of less-massive satellites^{5–10}. This formation process is thought to result in the multimodal optical colour distributions that are seen in the globular cluster systems of massive galaxies^{8,11,12}. Here we report optical observations of the massive relic-galaxy candidate NGC 1277—a nearby, un-evolved example of a high-redshift ‘red nugget’ galaxy^{13–17}. We find that the optical colour distribution of the cluster system of NGC 1277 is unimodal and entirely red. This finding is in strong contrast to other galaxies of similar and larger stellar mass, the cluster systems of which always exhibit (and are generally dominated by) blue clusters¹¹. We argue that the colour distribution of the cluster system of NGC 1277 indicates that the galaxy has undergone little (if any) mass accretion after its initial collapse, and use simulations of possible merger histories to show that the stellar mass due to accretion is probably at most ten per cent of the total stellar mass of the galaxy. These results confirm that NGC 1277 is a genuine relic galaxy and demonstrate that blue clusters constitute an accreted population in present-day massive galaxies.

We obtained imaging of NGC 1277 from the Advanced Camera for Surveys (in the g_{475W} and z_{850LP} filters) onboard the Hubble Space Telescope (HST/ACS) to characterize the globular cluster system of this galaxy (Extended Data Fig. 1). NGC 1277 has been identified as a massive (about $1.2 \times 10^{11} M_{\odot}$, where M_{\odot} is the mass of the Sun)¹⁵ relic galaxy on the basis of its high stellar-mass density, compactness, kinematics and old and metal-rich stellar populations^{14–17}. We postulate that if NGC 1277 is a true relic—that is, a nearby, un-evolved example of the ‘red nuggets’ seen at high redshift—then it should have accreted little or no stellar halo and should have few or no blue clusters.

The spatial distributions of cluster candidates from the HST/ACS imaging (Fig. 1) indicate that NGC 1277 and its companion galaxy in projection, NGC 1278, both have relatively rich cluster systems. Separating the clusters in the ACS field by their de-reddened colours by cutting at $(g_{475W} - z_{850LP})_0 = 1.15$ (the approximate separation between the red and blue peaks for a galaxy of this mass in the ACS Virgo Cluster Survey (VCS)¹¹), we find that NGC 1277 has a rich red cluster population but few blue clusters. This is in strong contrast to NGC 1278, which has a similar stellar mass to NGC 1277 (about $2.4 \times 10^{11} M_{\odot}$) but both red and blue clusters.

This visual impression is supported by looking at the colour distribution of the globular clusters in NGC 1277 compared with those in galaxies of similar stellar mass from the survey (Fig. 2). No obvious blue peak is seen in the colour distribution of clusters in NGC 1277, which

appears red with a tail to bluer colours. By contrast, the composite colour distribution of the similar-stellar-mass galaxies in the survey has both blue and red peaks. We ran Gaussian mixture modelling (GMM) tests (Methods) on these colour distributions, with a null hypothesis that the input colour distributions are unimodal. For NGC 1277, GMM cannot reject the null hypothesis for the globular clusters ($P = 0.371$, χ^2 test). However, under the assumption that two populations of clusters are present, GMM associates 99 clusters with a red population with a mean colour of $(g_{475W} - z_{850LP})_0 = 1.31 \pm 0.10$ (errors are 1σ unless otherwise stated) and 21 clusters with a metal-poor population with a mean colour of $(g_{475W} - z_{850LP})_0 = 0.86 \pm 0.24$. For the survey colour distribution, GMM rejects a unimodal distribution at high confidence ($P < 0.001$, χ^2 test), finding a roughly even split between blue (101 clusters) and red (110 clusters) clusters with colour peaks at $(g_{475W} - z_{850LP})_0 = 0.97 \pm 0.04$ and 1.36 ± 0.04 , respectively.

The results of the survey show that the cluster systems of present-day massive galaxies can have diverse colour distributions, but always have substantial numbers of blue clusters¹¹. For galaxies with stellar masses comparable to that of NGC 1277, the typical fraction of blue globular clusters is $f_{\text{blue}} = 0.50 \pm 0.10$ (ref. 11) or larger.

For NGC 1277, we find $f_{\text{blue}} = 0.17^{+0.12}_{-0.17}$ (systematic) $^{+0.06}_{-0.05}$ (random) (16th–84th percentiles; Fig. 3, Methods). This is an upper limit because the null hypothesis of unimodality is not rejected. NGC 1277 is clearly an outlier in that it has a significantly lower blue fraction than the survey galaxies with similar stellar masses (Fig. 3).

For massive galaxies in the survey, the blue cluster fractions are lower limits because the survey covered only the inner 1–2 effective radii (R_e) of these galaxies—a region where red clusters dominate over blue clusters owing to the more concentrated spatial distributions of the red clusters¹⁸. In the case of NGC 1277, which is very compact ($R_e = 1.2$ kpc) and located in the more distant Perseus cluster, we have probed the entire radial range (about $10R_e$) of the cluster system and are not biased against detecting blue clusters. Therefore, NGC 1277 truly lacks blue clusters when compared with normal galaxies (Fig. 3).

We rule out the possibility that a colour–magnitude relation in the blue globular clusters is driving the red-dominated peak of the clusters in NGC 1277. This relation is not seen in our data (Methods, Extended Data Fig. 6). In addition, our results are unaffected by possible non-linear colour–metallicity relations for the clusters^{19–21}. About 70% of the clusters in NGC 1277 have $(g_{475W} - z_{850LP})_0 > 1.1$, which is the approximate colour at which the empirical $(g_{475W} - z_{850LP})_0$ –metallicity relations for clusters in the Milky Way, M49 and M87 are linear¹¹.

NGC 1277 is roughly five times more compact than the majority of galaxies of the same stellar mass at $z = 0$ (ref. 15). As we have shown, NGC 1277 also lacks an obvious population of blue clusters, which are generally associated with a halo population in galaxies. The lack of a halo is consistent with the surface brightness profile of NGC 1277, which declines sharply beyond about 10 kpc from the centre of the galaxy owing to the presence of fewer stars (by a factor of about five) than in other galaxies of the same mass¹⁵. Assuming that NGC 1277

¹Instituto de Astrofísica de Canarias, Calle Vía Láctea, La Laguna, Spain. ²University of La Laguna, Avenida Astrofísico Francisco. Sánchez, La Laguna, Spain. ³Max-Planck Institut für Astronomie, Königstuhl 17, D-69117, Heidelberg, Germany. ⁴Department of Astronomy, Yale University, 06511 New Haven, Connecticut, USA.

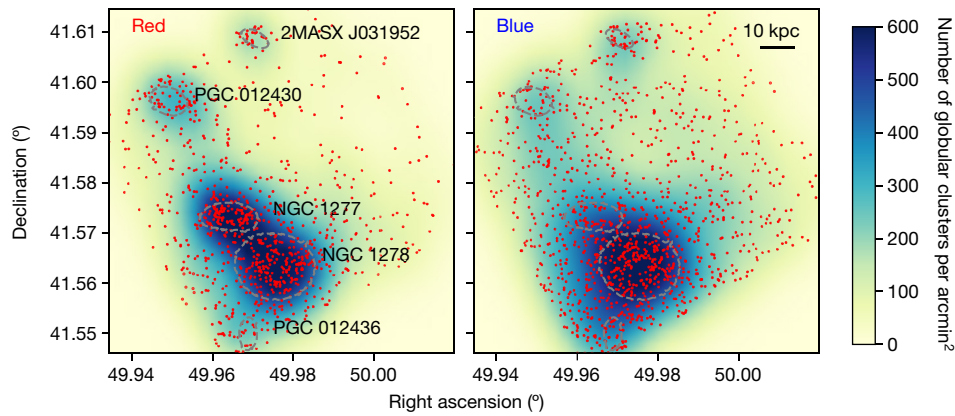


Figure 1 | Spatial distribution of clusters in the HST/ACS field. North is up, east is to the right. Individual clusters are shown as red points. The locations of the named galaxies are indicated (grey dashed ellipsoids) by isophotes corresponding to $23 \text{ mag arcsec}^{-2}$ in g_{475W} . NGC 1277 is left of centre in the plots, with the neighbouring galaxy NGC 1278 located around 50 arcsec (17 kpc) to the southeast in projection. The red (left) and blue (right) clusters have been separated by taking a cut

at $(g_{475W} - z_{850LP})_0 = 1.15$, typical of the peak separation between the red and blue clusters in galaxies of this stellar mass¹¹. The colour scale shows Gaussian kernel density maps of the globular clusters, constructed with a kernel with a width of 10 arcsec. Galaxies in the field can clearly be identified by the relative prevalence of clusters. NGC 1278 has both red and blue cluster populations, whereas NGC 1277 has a substantial population only of red clusters.

obeys the same size–stellar-mass relation as that of quiescent galaxies of the same stellar mass at $z = 2$, our findings suggest either that it has been stripped of its outer envelope and clusters and reduced in size, or that it failed to grow substantially in size and in the number of blue clusters since $z = 2$. We have explored both scenarios and favour the latter, that NGC 1277 is a genuine relic galaxy.

The tidal radius of NGC 1277—at which stars and clusters become unbound from the galaxy—can be estimated analytically²². In the extreme case that NGC 1277 is currently at pericentre within the Perseus cluster and that its clusters are on prograde orbits with respect to its orbit within Perseus, we calculate a lower limit for the tidal radius of 11 kpc. This lower limit corresponds to roughly $2R_e - 3R_e$ in galaxies with similar stellar masses to that of NGC 1277 and is a radius at which at least 50% the clusters in the population should be blue¹¹. This result suggests that stripping has not affected the physical size or the cluster system of NGC 1277. Furthermore, the central stellar density of NGC 1277 is a factor of 2–3 higher than that of the inner 1 kpc of galaxies of similar stellar mass¹⁵. Tidal stripping alone cannot increase the central stellar density²², which implies that NGC 1277 was born

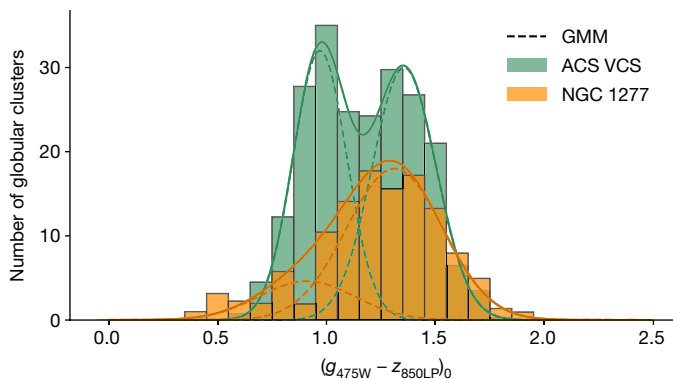


Figure 2 | The colour distribution of clusters in NGC 1277 compared with that of the composite cluster system. The colour distribution for NGC 1277 (orange histogram) is constructed over the full radial extent of its cluster system (11 kpc) and has been corrected for the contamination from NGC 1278 clusters; in projection, NGC 1278 is 17 kpc from NGC 1277. The composite cluster system (ACS VCS; green histogram) is constructed from four galaxies¹¹ with stellar masses similar to that of NGC 1277 (roughly $1.2 \times 10^{11} M_\odot$). The dashed curves indicate the Gaussian components of the cluster systems obtained from GMM²⁸; the solid curves show the sum of these components. The composite cluster system is clearly bimodal, whereas NGC 1277 lacks an obvious blue peak.

both dense and compact. This raises the question of why NGC 1277 should have suffered severe tidal stripping while larger galaxies of similar mass in the same environment—such as NGC 1278—have not. Finally, there is no evidence for tidal tails or streams in the outer region of NGC 1277¹⁵, or for stripping in the stellar orbits, rotation curve or mass distribution of NGC 1277¹⁷.

By contrast, NGC 1277 strongly resembles the low-redshift equivalent of higher-redshift compact systems. It lies on the size–stellar-mass relation of $z = 2$ galaxies^{15,17} and has a high central stellar density that is consistent with massive galaxies at these redshifts²³. Schwarzschild dynamical modelling²⁴ shows that NGC 1277 is dominated by high-angular-momentum orbits, which implies that it formed in a largely dissipative process. Previous results^{17,24} suggest that NGC 1277 is essentially a massive disk with no classical, pressure-supported component.

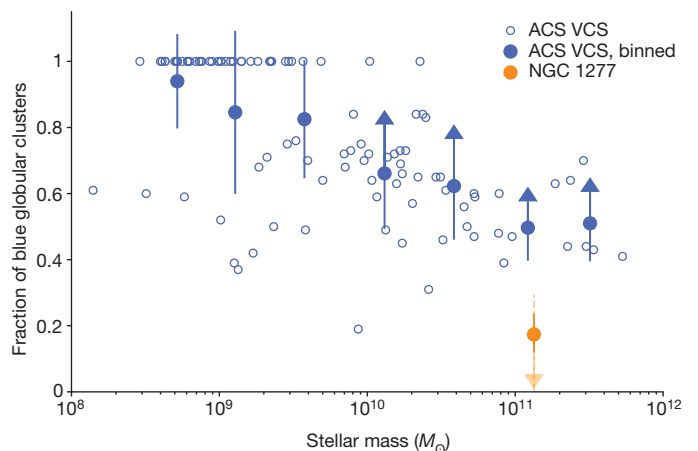


Figure 3 | The fraction of blue clusters in galaxies of a given stellar mass. Data for the survey galaxies (ACS VCS; open blue circles) come from ref. 11. NGC 1277 (filled orange circle) has a very small fraction of blue clusters when compared with other galaxies of similar stellar mass. For the binned survey data (filled blue circles), the 16th–84th percentiles of the distributions are shown as error bars; the blue fractions for the brightest survey galaxies are lower limits (indicated by the arrowheads) because only the central parts of the galaxy cluster systems were observed. The systematic uncertainty on the blue fraction for NGC 1277 is shown as the faint, dashed error bar (16th–84th percentiles); the random uncertainties from Monte Carlo simulations are smaller and shown as the solid error bar (16th–84th percentiles). The blue fraction for the clusters in NGC 1277 is an upper limit (indicated by the arrowhead) because the null hypothesis of unimodality cannot be rejected on the basis of observations.

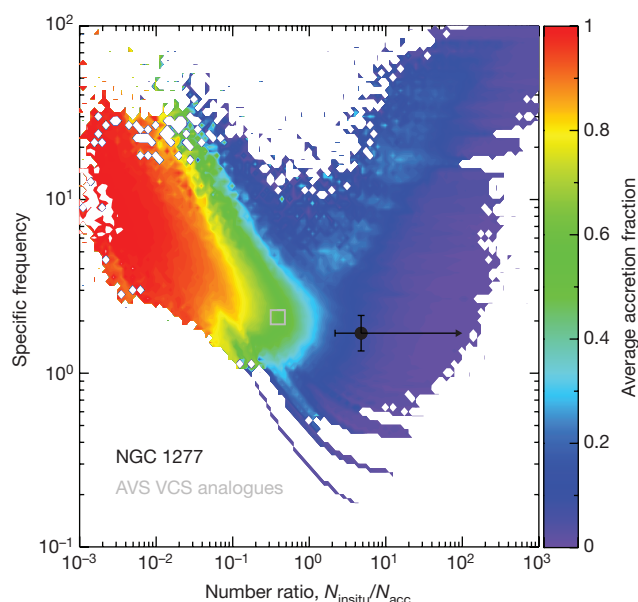


Figure 4 | Predictions from the simulated merger histories of NGC 1277. The total specific frequency—the number of clusters per unit galaxy luminosity—is shown against the number ratio of clusters formed *in situ* (N_{insitu}) and those accreted (N_{acc}) in the simulated merger histories. The colours represent the average accretion fraction for each merger history. The constraints for NGC 1277 (black circle, with error bars indicating the 16th–84th percentiles, the arrow indicates that our measurements are an upper limit for the accreted fraction) that we determined from the observations are often satisfied in histories with low accretion fractions (about 10%). Merger histories with higher accretion fractions rarely lead to the colour ratio that we observed without overproducing the total number of clusters. By contrast, similar-mass galaxies in the ACS VCS (grey box) are well reproduced by merger histories with accretion fractions of around 50%.

These kinematic properties are remarkably similar to those of the lensed ‘dead disk’ at $z \approx 2.1$ (ref. 25), which has similar stellar mass to NGC 1277 and is similarly compact ($R_e = 1.73^{+0.34}_{-0.27}$ kpc). In addition, the old, metal-rich stellar population of NGC 1277 is similar to that of other nearby relic-galaxy candidates¹⁷, to that expected for passively evolving $z = 2$ red nuggets and to that predicted for the redshift evolution of massive, compact systems in cosmological simulations^{26,27}. Finally, the fact that NGC 1277 is in a massive galaxy cluster is not unexpected, observationally or theoretically (Methods).

Our findings for the cluster system of NGC 1277 provide further constraints on the evolutionary histories of relic galaxies. In accretion models for the formation of the blue cluster population in massive galaxies, the location of the blue peak and the relative fraction of blue clusters provides insight into the total mass and the properties of accreted satellites of any given galaxy^{6,8,9}. We ran a series of analytic merger models based on our observations (Fig. 4, Methods, Extended Data Figs 7, 8) to constrain the accreted mass fraction of NGC 1277. From this modelling we find that the observed blue cluster fraction of NGC 1277 is characteristic of galaxies that have accreted about 10% or less of their stellar mass, with the rest of the stars having formed *in situ*. By contrast, mass-matched ACS VCS galaxies have accretion fractions of 50%–90%. We therefore find that the blue fraction of globular clusters approximately traces the total fraction of accreted mass in galaxies, and that NGC 1277 has accreted very little mass (if any) over its lifetime compared with other galaxies of similar stellar mass.

We conclude that NGC 1277 provides a benchmark for the detailed study of the properties of massive galaxies that have not evolved since $z = 2$. High-resolution imaging of the cluster systems of more relic galaxies will enable the reconstruction of their mass-accretion histories.

Online Content Methods, along with any additional Extended Data display items and Source Data, are available in the online version of the paper; references unique to these sections appear only in the online paper.

Received 5 July 2017; accepted 3 January 2018.

Published online 12 March 2018.

- Khochfar, S. & Silk, J. A simple model for the size evolution of elliptical galaxies. *Astrophys. J.* **648**, L21–L24 (2006).
- Oser, L., Naab, T., Ostriker, J. P. & Johansson, P. H. The cosmological size and velocity dispersion evolution of massive early-type galaxies. *Astrophys. J.* **744**, 63 (2012).
- Ceverino, D., Dekel, A., Tweed, D. & Primack, J. Early formation of massive, compact, spheroidal galaxies with classical profiles by violent disc instability or mergers. *Mon. Not. R. Astron. Soc.* **447**, 3291–3310 (2015).
- van Dokkum, P. G. et al. The growth of massive galaxies since $z=2$. *Astrophys. J.* **709**, 1018–1041 (2010).
- Searle, L. & Zinn, R. Compositions of halo clusters and the formation of the galactic halo. *Astrophys. J.* **225**, 357–379 (1978).
- Côté, P., Marzke, R. O. & West, M. J. The formation of giant elliptical galaxies and their globular cluster systems. *Astrophys. J.* **501**, 554–570 (1998).
- Beasley, M. A., Baugh, C. M., Forbes, D. A., Sharples, R. M. & Frenk, C. S. On the formation of globular cluster systems in a hierarchical Universe. *Mon. Not. R. Astron. Soc.* **333**, 383–399 (2002).
- Tonini, C. The metallicity bimodality of globular cluster systems: a test of galaxy assembly and of the evolution of the galaxy mass-metallicity relation. *Astrophys. J.* **762**, 39 (2013).
- Leaman, R., VandenBerg, D. A. & Mendel, J. T. The bifurcated age–metallicity relation of Milky Way globular clusters and its implications for the accretion history of the galaxy. *Mon. Not. R. Astron. Soc.* **436**, 122–135 (2013).
- Kruijssen, J. M. D. Globular clusters as the relics of regular star formation in normal high-redshift galaxies. *Mon. Not. R. Astron. Soc.* **454**, 1658–1686 (2015).
- Peng, E. W. et al. The ACS Virgo Cluster Survey. IX. The color distributions of globular cluster systems in early-type galaxies. *Astrophys. J.* **639**, 95–119 (2006).
- Brodie, J. P. & Strader, J. Extragalactic globular clusters and galaxy formation. *Annu. Rev. Astron. Astrophys.* **44**, 193–267 (2006).
- Damjanov, I. et al. Red nuggets at $z=1.5$. Compact passive galaxies and the formation of the Kormendy relation. *Astrophys. J.* **695**, 101–115 (2009).
- van den Bosch, R. C. E. et al. An over-massive black hole in the compact lenticular galaxy NGC 1277. *Nature* **491**, 729–731 (2012).
- Trujillo, I., Ferré-Mateu, A., Balcells, M., Vazdekis, A. & Sánchez-Blázquez, P. NGC 1277: a massive compact relic galaxy in the nearby Universe. *Astrophys. J.* **780**, L20 (2014).
- Martín-Navarro, I. et al. The initial mass function of a massive relic galaxy. *Mon. Not. R. Astron. Soc.* **451**, 1081–1089 (2015).
- Yildirim, A. et al. The structural and dynamical properties of compact elliptical galaxies. *Mon. Not. R. Astron. Soc.* **468**, 4216–4245 (2017).
- Hargis, J. R. & Rhode, K. L. Globular cluster systems and their host galaxies: comparison of spatial distributions and colors. *Astrophys. J.* **796**, 62 (2014).
- Blakeslee, J. P. et al. Optical and infrared photometry of globular clusters in NGC 1399: evidence for color–metallicity nonlinearity. *Astrophys. J.* **746**, 88 (2012).
- Yoon, S.-J. et al. Nonlinear color–metallicity relations of globular clusters. III. On the discrepancy in metallicity between globular cluster systems and their parent elliptical galaxies. *Astrophys. J.* **743**, 150 (2011).
- Chies-Santos, A. L. et al. An optical/NIR survey of globular clusters in early-type galaxies. III. On the colour bimodality of globular cluster systems. *Astron. Astrophys.* **539**, A54 (2012).
- Read, J. I., Wilkinson, M. I., Evans, N. W., Gilmore, G. & Kleyna, J. T. The tidal stripping of satellites. *Mon. Not. R. Astron. Soc.* **366**, 429–437 (2006).
- Szomoru, D., Franx, M. & van Dokkum, P. G. Sizes and surface brightness profiles of quiescent galaxies at $z=2$. *Astrophys. J.* **749**, 121 (2012).
- Yildirim, A. et al. MRK 1216 and NGC 1277 – an orbit-based dynamical analysis of compact, high-velocity dispersion galaxies. *Mon. Not. R. Astron. Soc.* **452**, 1792–1816 (2015).
- Toft, S. et al. A massive, dead disk galaxy in the early Universe. *Nature* **546**, 510–513 (2017).
- Wellons, S. et al. The diverse evolutionary paths of simulated high- z massive, compact galaxies to $z=0$. *Mon. Not. R. Astron. Soc.* **456**, 1030–1048 (2016).
- Furlong, M. et al. Size evolution of normal and compact galaxies in the EAGLE simulation. *Mon. Not. R. Astron. Soc.* **465**, 722–738 (2017).
- Muratov, A. L. & Gnedin, O. Y. Modeling the metallicity distribution of globular clusters. *Astrophys. J.* **718**, 1266–1288 (2010).

Acknowledgements We thank C. dalla Vecchia, J. Sanchez Almeida, C. Brook, S. Wellons, M. Fouesneau, A. Vazdekis, B. Dullo, J. Read and G. van de Ven for discussions, and J. Roman and A. Serrano Borlaf for assistance with image alignment. M.A.B. and I.T. acknowledge support from grant AYA2016-77237-C3-1-P from the Spanish Ministry of Economy and Competitiveness (MINECO). R.L. acknowledges funding from a Natural Sciences and Engineering Research Council of Canada PDF award. This research has made use of the NASA/IPAC Extragalactic Database (NED), which is operated by the Jet Propulsion Laboratory, California Institute of Technology, under contract with NASA, and is based on observations made with the NASA/ESA Hubble Space Telescope,

which is operated by the Association of Universities for Research in Astronomy under NASA contract NAS 5-26555. These observations are associated with programme GO-14215. Support for this work was provided by NASA through grant HST-GO-4215 from the Space Telescope Science Institute, operated by AURA under NASA contract NAS 5-26555. This research has made use of NASA's Astrophysics Data System and extensive use of Python and Scipy.

Author Contributions M.A.B. led the data processing and analysis, contributed to the interpretation and HST proposal preparation and produced Figs 1–3 and Extended Data Figs 2–6. I.T. contributed to the analysis and the interpretation, produced Extended Data Fig. 1 and lead the HST proposal preparation. R.L. generated and analysed the analytic merger models, produced Fig. 4 and Extended Data Figs 7, 8, and contributed to the analysis, interpretation and

HST proposal preparation. M.M. contributed to the analysis, interpretation and HST proposal preparation. All authors contributed to the overall design of this project.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Publisher's note: Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations. Correspondence and requests for materials should be addressed to M.A.B. (beasley@iac.es).

Reviewer Information *Nature* thanks R. Abraham, K. Glazebrook and L. Spitler for their contribution to the peer review of this work.

METHODS

HST/ACS photometry. We obtained HST/ACS imaging of NGC 1277 in F475W (g_{475W}) and F850LP (z_{850LP}) (two orbits; GO programme ID, 14215; Extended Data Fig. 1). The total exposure time was 2,280 s (F475W) and 2,432 s (F850LP). We adopt a standard cosmological model with $H_0 = 70 \text{ km s}^{-1} \text{ Mpc}^{-1}$, $\Omega_m = 0.3$ and $\Omega_\Lambda = 0.7$. The redshift assumed here for NGC 1277, $z = 0.0169$, corresponds to a galaxy distance of 73.3 Mpc ($(m - M)_0 = 34.33$) and a spatial scale of $344 \text{ pc arcsec}^{-1}$.

The HST data were pipeline-reduced, including correction for charge-transfer efficiency. We calculated zero-points for the AB magnitude system using the file header information. Aperture and point spread function (PSF) photometry was performed on the ACS imaging using Source Extractor²⁹ with the add-on PSFEx³⁰. Our photometry was corrected for foreground extinction using line-of-sight reddening estimates³¹. Source Extractor was run in dual-image mode using un-sharp masks as detection images with photometry performed on the original images. We selected matched sources in the g_{475W} and z_{850LP} filters. Aperture magnitudes were measured using a range of aperture radii from 3 to 50 pixels. We applied aperture corrections to the three-pixel aperture magnitudes by correcting to ten-pixel (0.5 arcsec) apertures based on bright, isolated stars, and then correcting to infinity using the corrections tabulated in ref. 32. For artificial point-source tests, we used PSFEx to construct PSFs based on bright, isolated stars across the ACS field.

Completeness tests. Completeness tests were performed by injecting artificial point sources into the images (using the PSFEx PSFs) and determining the recovery fraction as a function of magnitude with Source Extractor. In regions approximately 20 arcsec from bright galaxies, we are 100% complete to $g_{475W} = 27.3 \text{ mag}$ and $z_{850LP} = 26.7 \text{ mag}$, with typical colour uncertainties of $\delta(g_{475W} - z_{850LP})_0 = 0.08 \text{ mag}$. We take this as a conservative photometric limiting magnitude. However, the completeness varies as a function of position and therefore we also calculated photometric completeness by repeating our tests as a function of position and magnitude across the field. In the very central regions (less than 1.5 arcsec; 0.5 kpc) of NGC 1277 and the neighbouring NGC 1278, we recover few clusters owing to the high surface brightness of the galaxies. Photometric errors were characterized by measuring the difference between the input and output magnitudes of 50,000 artificial point sources placed randomly across the images.

To separate point sources (globular cluster candidates) from extended sources, we obtained the Source Extractor output *flux_radius* as a function of magnitude for our artificial point sources. All real sources identified within the region defined by the artificial sources were regarded as cluster candidates. In total, to $g_{475W} = 27.3 \text{ mag}$, we detect 2,286 objects in the magnitude range consistent with being clusters. Colour–magnitude diagrams for all of the cluster candidates, and for cluster candidates around NGC 1278, NGC 1277 and an example background field are shown in Extended Data Fig. 2.

Characterizing the colour distributions of the cluster systems. The spatial density distribution of candidate clusters (Fig. 1) was constructed directly from the photometry, after masking a small region (radius of 3 arcsec) centred on an uncatalogued dwarf located about 16 arcsec to the west of NGC 1277 (right ascension, 03 h 19 min 50.1 s; declination, $+41^\circ 34' 22.10''$). A kernel-density-estimate map was constructed using the Python SciPy routine *gaussian_kde* with the Gaussian kernel determined using ‘Scott’s rule’³³. As expected, the globular clusters aggregate around known galaxies (Fig. 1). However, taking a colour cut at $(g_{475W} - z_{850LP})_0 = 1.15$ —the approximate separation between the red and blue peaks for a galaxy the mass of NGC 1277 in ref. 11—NGC 1277 seems to ‘drop out’ of the blue-cluster-density map. This is in contrast with NGC 1278, which has both red and blue clusters.

To create the colour distribution of NGC 1277 (Fig. 2), we had to take into account interloping clusters from the neighbouring NGC 1278 and contributions from intracluster clusters. We constructed a ‘master background’ by selecting five regions located at a distance from NGC 1278 equal to the separation between NGC 1277 and NGC 1278, and sufficiently distant from NGC 1277 so as to not overlap with its cluster system (which has an extent of about 11 kpc; Extended Data Fig. 3). The colour distribution of NGC 1277 clusters was built by selecting all cluster candidates within 11 kpc (about $10R_e$) and subtracting the master background normalized to the area covered by the cluster system of NGC 1277. This process also removes intracluster clusters from the cluster system of NGC 1277. The colour distribution of this master background is shown in Extended Data Fig. 4. The composite cluster system in Fig. 2 was constructed by selecting galaxies from ref. 11 that bracket the stellar mass of NGC 1277, giving a mean stellar mass of about $1.1 \times 10^{11} M_\odot$ with a dispersion of $0.4 \times 10^{11} M_\odot$. To compare with NGC 1277, we took a magnitude cut in the survey data at $g_{475W} = 23.98$ ($(m - M)_0 = 31.01$), which corresponds to our photometric depth.

To explore the subpopulations in the colour distributions, we ran a GMM code²⁸ on the binned colour distributions. For the visibly bimodal colour distribution of the survey composite, in the heteroscedastic (different variances

between populations) case we obtain means of $(g_{475W} - z_{850LP})_0 = 0.97 \pm 0.04$ and 1.36 ± 0.04 with Gaussian full-widths at half-maximum (FWHMs) of 0.12 ± 0.02 and 0.15 ± 0.02 , respectively. 101 clusters are associated with the blue peak and 110 clusters with the red peak. A unimodal distribution is rejected at high confidence ($P < 0.001$, χ^2 test). Similar results were obtained in the homoscedastic case.

For NGC 1277, a unimodal distribution cannot be rejected because the null hypothesis has $P = 0.371$ confidence (χ^2 test). Notwithstanding the possibility of unimodality, in the homoscedastic case for two populations we obtain means of $(g_{475W} - z_{850LP})_0 = 0.86 \pm 0.24$ and 1.31 ± 0.10 each with a FWHM of 0.21 ± 0.08 . 21 clusters are associated with the blue peak and 99 clusters with the red peak (Fig. 2). We define a blue fraction $f_{\text{blue}} = N_{\text{blue}} / (N_{\text{blue}} + N_{\text{red}})$.

We consider two sources of uncertainty for f_{blue} . The uncertainty returned from GMM on the number of blue and red clusters we regard as our systematic uncertainty. In addition, a random uncertainty on f_{blue} comes from our background subtraction. To quantify the true value of f_{blue} and this uncertainty, we performed 1,000 Monte Carlo simulations in which we selected clusters randomly from our background regions (until the observed background level was reached), and varied the radial apertures of the background and NGC 1277 selection regions by $\pm 3 \text{ kpc}$ ($\pm 9 \text{ arcsec}$). From this we built new colour distributions and ran GMM on these colour distributions to obtain f_{blue} (Extended Data Fig. 5). For the homoscedastic case, we obtain $f_{\text{blue}} = 0.18^{+0.12}_{-0.17}$ (systematic) ± 0.04 (random) (16th–84th percentiles); for the heteroscedastic case, we obtain $f_{\text{blue}} = 0.17^{+0.12}_{-0.17}$ (systematic) $^{+0.06}_{-0.05}$ (random) (16th–84th percentiles). We conservatively take this latter value as the true blue fraction and its associated uncertainties. From the Monte Carlo simulations, we find $N_{\text{blue}}/N_{\text{red}} = 0.21^{+0.25}_{-0.20}$ (16th–84th percentiles). We consider these as upper limits because the unimodal hypothesis cannot be rejected.

As a sanity check, we compared the colour distributions of NGC 1277 with those of NGC 1278 (Extended Data Fig. 4). We remove contributions from intracluster clusters to the colour distribution of NGC 1278 by selecting clusters in a region away from bright galaxies (centred on a right ascension of 3 h 19 min 57.0 s and declination of $+41^\circ 35' 30.0''$) over an area normalized to that of our cluster-selection region. NGC 1278 shows a prominent blue peak of clusters and a less prominent red peak. GMM (heteroscedastic case) locates peaks at $(g_{475W} - z_{850LP})_0 = 0.95 \pm 0.24$ and 1.40 ± 0.10 . These solutions are in excellent agreement with expectations¹¹.

Surface-density profiles of the clusters. The surface-density profile of clusters (Extended Data Fig. 3) was constructed by counting globular clusters in semi-circular annuli and dividing by the area of each semi-annulus. We only counted clusters in the northern half of the galaxy—bisected by the major axis of the galaxy—to minimize contamination from the cluster system of NGC 1278. The inner two radial bins have been corrected on the basis of our completeness tests. The clusters closely follow the galaxy light¹⁵. This is a characteristic property of red clusters in massive galaxies³⁴. By contrast, blue clusters generally have spatial distributions that are more extended than the galaxy stars^{18,34}. To define the radial extent of the system, and to locate the background level, we fitted the data with a modified Sérsic function³⁵:

$$N_{\text{GC}}(R) = N_e \exp \left[-b_n \left[\left(\frac{R}{R_e} \right)^{1/n} - 1 \right] \right] + \text{bkg}$$

with N_e the surface density of clusters at radius R_e , n the Sérsic index and bkg the background value; b_n is related to n as $b_n = 1.9992n - 0.3271$.

A Sérsic fit (with $N_e = 7.8 \pm 0.7 \text{ kpc}^{-2}$, $R_e = 2.6 \pm 0.2 \text{ kpc}$, $n = 0.9 \pm 0.3$, $\text{bkg} = 0.27 \pm 0.01 \text{ kpc}^{-2}$) to the surface-density profile of the clusters (Extended Data Fig. 3), along with visual inspection, indicates that we reach the background level about 11 kpc from the centre of NGC 1277. We consider this the full radial extent of the NGC 1277 cluster system.

Globular-cluster luminosity functions. We constructed cluster luminosity functions for the clusters in NGC 1277 to obtain an independent distance estimate to the galaxy using the turn-over of the luminosity function of the clusters as a standard candle³⁶, and to calculate the total size of the cluster system. Again, we selected only from the northern half of the galaxy to minimize interlopers from NGC 1278. We binned the g_{475W} and z_{850LP} magnitudes of the clusters as a function of magnitude (0.5-mag bins) and ran a version of a maximum likelihood code³⁷ that fits for the mean (μ), FWHM (σ) and normalization of the distribution. The code takes into full account the background, incompleteness and photometric errors. In g_{475W} (z_{850LP}), we obtain $\mu = 26.95 \pm 0.15 \text{ mag}$ ($25.80 \pm 0.20 \text{ mag}$) and $\sigma = 1.25 \pm 0.15$ (1.35 ± 0.20). Assuming ‘universal’ absolute magnitudes for the turn-over of the luminosity function of $M_{g_{475W}} = -7.2 \text{ mag}$ ($M_{z_{850LP}} = -8.4 \text{ mag}$)³⁸, we obtain distance moduli of $(m - M)_0 = 34.15 \pm 0.15$ ($(m - M)_0 = 34.20 \pm 0.20$). These values are in excellent agreement with our adopted distance to NGC 1277.

Total cluster population of NGC 1277. We counted the total number of clusters brighter than the turn-over of the luminosity function of the clusters (26.95 ± 0.15 mag in g_{475W}) that lie within 11 kpc of the galaxy centre within the northern half of NGC 1277. In so doing, we detect 92 ± 18 clusters to our photometric limit, where the uncertainties come from the uncertainty of ± 0.2 mag in the peak position of the luminosity function. From this number we subtracted the expected contribution from NGC 1278 clusters (27 ± 8 clusters). Correcting for radial incompleteness, this number becomes 93.9 ± 23.0 . We then doubled the total number of clusters to account for the undetected faint half of the luminosity function, and doubled it again because we constructed the surface-density profile in the northern half of the galaxy. We arrive at a total cluster population of $N_{\text{cluster}} = 376 \pm 94$. For $M_V = -20.87$ (NASA Extragalactic Database, <https://ned.ipac.caltech.edu/>), this cluster population yields a specific frequency of $S_N = 1.7 \pm 0.4$, consistent with galaxies in the stellar mass range of NGC 1277, which typically have $S_N \approx 2.0$ (ref. 39).

The blue tilt. The ‘blue tilt’ manifests as a colour–magnitude relation in the blue clusters whereby the brightest blue clusters become redder with increasing luminosity^{40–42}. This can make the colour distributions for the brightest clusters look unimodal. We investigated this issue by making colour distributions from the median bin values of our Monte Carlo simulations for the NGC 1277 clusters binned by magnitude (Extended Data Fig. 6). The mean colours of all three bins are very similar and predominantly red, with $\langle g_{475W} - z_{850LP} \rangle = 1.22$ ($22.0 < z_{850LP} < 24.0$), $\langle g_{475W} - z_{850LP} \rangle = 1.19$ ($24.0 < z_{850LP} < 25.0$) and $\langle g_{475W} - z_{850LP} \rangle = 1.23$ ($25.0 < z_{850LP} < 26.5$). This behaviour is not expected if a blue tilt were driving the observed colour distributions. The majority of the clusters with $\langle g_{475W} - z_{850LP} \rangle \leq 1.1$ are consistent with the background residuals. However, a few of the brightest clusters (eight) are above the background. These ‘blue’ clusters constitute a small fraction of the overall cluster population.

The environments of relic galaxies. NGC 1277 lies 3.8 arcmin (roughly 80 kpc) in projection to the north of the central massive galaxy NGC 1275 in the Perseus cluster of galaxies. However, it is not unexpected that such a relic galaxy should be found in such a dense environment.

Observational studies that have looked specifically at the preferred environments of massive compact galaxies in the nearby Universe find that the fraction of present-day compact, quiescent galaxies with stellar masses of more than about $3 \times 10^{10} M_\odot$ that lie in low-density environments is approximately 4.4%⁴³. The corresponding fraction in nearby galaxy clusters is about 22%⁴⁴. Similar dependencies on environmental density have been obtained from different samples^{45,46}. Of the 16 best-studied nearby, massive, compact galaxies (including NGC 1277), 7 are in clusters, 4 in groups and 5 in isolated systems¹⁷.

On the theoretical side, dark-matter-only simulations have been used⁴⁷ to explore the clustering properties of massive, compact galaxies, and reveal that the fraction of massive, compact galaxies is about five times larger in the most massive structures than in low-mass dark-matter haloes. Similarly, semi-analytic models have been used⁴⁵ to explore the environments of compact, massive galaxies, and reveal that these galaxies represent about 0.04% of the total galaxy population, but about 0.18% of the galaxy population in clusters. In addition, compact galaxies preferentially lie in the centres of clusters (the inner 0.2 virial radii, which corresponds to 400 kpc in Perseus), whereas non-compact galaxies of the same mass have mean cluster-centric radii of around 0.4 virial radii⁴⁵. The evolution of massive compact systems at $z = 2$ has previously been explored²⁶ in cosmological, hydrodynamical simulations: of a sample of 35 galaxies, roughly 30% remained sufficiently undisturbed to be defined as relics at $z = 0$. Of these relics, about 20% survive as satellites in galaxy clusters.

In summary, although massive, compact, ‘relic’ galaxies are rare in the nearby Universe (with space densities of 10^{-5} Mpc^{-3})^{48,49}, they are expected to be preferentially found in the most massive structures, such as the Perseus cluster of galaxies.

Accretion models for cluster formation. To obtain a quantitative understanding of the link between the colour distributions of clusters and the accretion history of a galaxy, we constructed a library of analytic merger histories for NGC 1277. We ran 1.5×10^6 model realizations of NGC 1277 in which we reconstruct its total present-day stellar mass (M_*) by assuming that it had accreted some fraction of its stellar mass $0 \leq f_{\text{acc}} \leq 1$. To test rare, or non-cosmological merger histories, for each realization we draw satellite galaxies from a sub-halo dark-matter mass function with a randomly selected slope $-1 \leq \alpha \leq 0$. These satellites are assigned stellar mass stochastically via empirical abundance-matching relations⁵⁰ and their stellar mass is added to the *in situ* stellar mass of NGC 1277 ($M_{\text{insitu}} = (1 - f_{\text{acc}})M_{*,\text{obs}}$) until the observed total present-day stellar mass of NGC 1277 ($M_{*,\text{obs}}$) is reached for that merger-history realization. This necessitates a dynamic sub-halo mass-function cut-off, because the maximum accreted sub-halo at any point cannot be more than the total accreted mass for that trial and must be less than the *in situ* mass. This ensures a working definition of the primary being most massive and, importantly, preserves an accretion fraction of f_{acc} and total mass of $M_{*,\text{obs}}$ at the end of each trial.

The stellar mass of each sub-halo sets its metallicity^{9,51} (and the metallicity of the clusters that are accreted with it), and the dark-matter mass of the satellite provides the specific frequency of clusters for that satellite^{39,52} (with stochastic sampling of the scatter in all of the relations). The same relations are used to construct the cluster population of the *in situ* component of NGC 1277 in every realization. Finally, the cluster metallicities are converted to $g_{475W} - z_{850LP}$ colours¹¹.

This exercise is qualitatively similar to that used previously⁸ with the added benefit that we allow for a flexible slope of the sub-halo mass function, rather than constraining it on the basis of average accretion histories in cold dark matter simulations. This addition allows us to explore how the cluster colour distributions and galaxy stellar mass could arise in potentially rare scenarios with either very low or high numbers of mergers or consecutive extreme-mass-ratio mergers. Although simple, these models recover the ensemble properties of the observed cluster colour and number distributions for galaxies of the mass of NGC 1277. Extended Data Fig. 7 shows the accreted fraction of each realization of NGC 1277 (with each dot representing one possible merger history that reproduces the total mass of NGC 1277) versus the ratio of the number of *in situ* to accreted (red to blue) clusters for NGC 1277 in that realization. We have conservatively used the minimum $N_{\text{insitu}}/N_{\text{acc}}$ value allowed within the error bars ($N_{\text{insitu}}/N_{\text{acc}} = 2.17$) rather than $N_{\text{insitu}}/N_{\text{acc}} = N_{\text{red}}/N_{\text{blue}} = 99/21 = 4.76$. The observed colour ratio of the clusters in NGC 1277 is characteristic of galaxies that have undergone merger histories resulting in the host accreting 12% of the present-day stellar mass. In comparison, galaxies of comparable mass to NGC 1277 in the ACS VCS sample have observed colour distributions ($N_{\text{red}}/N_{\text{blue}} \leq 1$) typical of having accreted 50%–90% of their stellar mass. A very small percentage (about 0.02%) of these high- f_{acc} models can produce cluster colour ratios similar to that for NGC 1277. However, Fig. 4 and Extended Data Fig. 8 show that these high- f_{acc} assembly histories tend to over-predict the total number of clusters per unit mass ($S_N \geq 2$) and to result in systems with dark-matter masses of $10^{14} M_\odot$ or more, which is ruled out for NGC 1277 by dynamical modelling^{17,24}. Extended Data Fig. 8 illustrates that the typical merger histories in our models that successfully reproduce the S_N and colour distribution of NGC 1277 tend to have accreted only $f_{\text{acc}} = 12\% \pm 8\%$ of their stellar mass, to have had a merger event with a mass ratio of at most 1:10 and to reside in under-massive dark-matter haloes ($M_*/M_{\text{DM}} \approx 0.06$) when compared with galaxies of the same stellar mass as NGC 1277. These models support the idea that NGC 1277 has not accreted substantial amounts of dark-matter-rich sub-haloes, leaving it deficient both in dark-matter mass and in blue, accreted clusters.

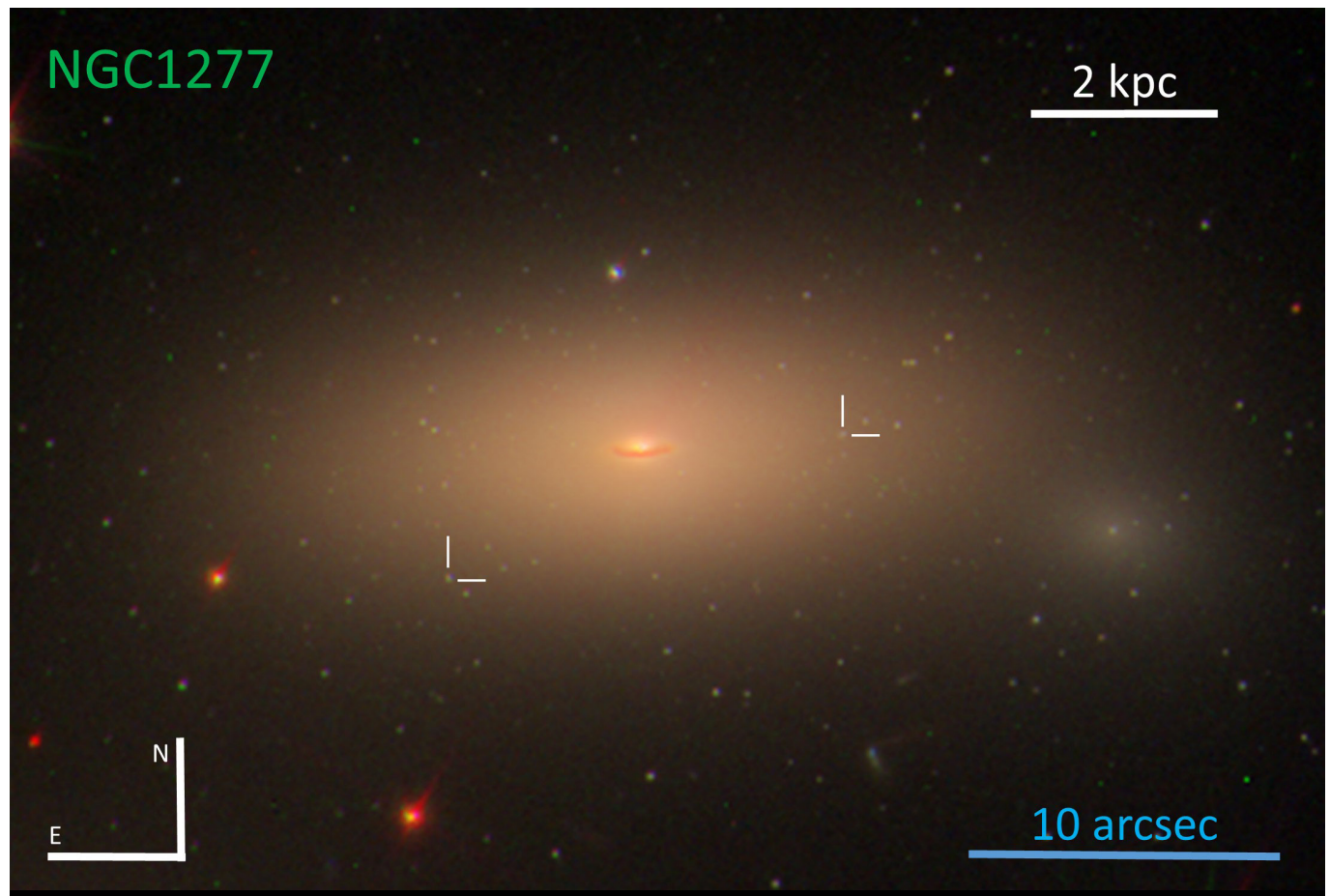
Sample size. No statistical methods were used to predetermine sample size.

Code availability. The photometry software Source Extractor and PSFEx are publicly available at <https://www.astromatic.net/software/sextractor> and <https://www.astromatic.net/software/psfex>. The GMM code is publicly available at <http://www-personal.umich.edu/~ognedin/gmm/>. A version of the globular cluster luminosity function fitting code is available on request. The code used for the modelling of the accretion histories of NGC 1277 will be made available in the future.

Data availability. The HST data used was obtained under GO programmes 14215 and 10546 and are publicly available at https://archive.stsci.edu/proposal_search.php?id=14215&mission=hst and https://archive.stsci.edu/proposal_search.php?id=10546&mission=hst. Data products such as the point-source photometry and results of the Monte Carlo simulations are available from the corresponding author on request.

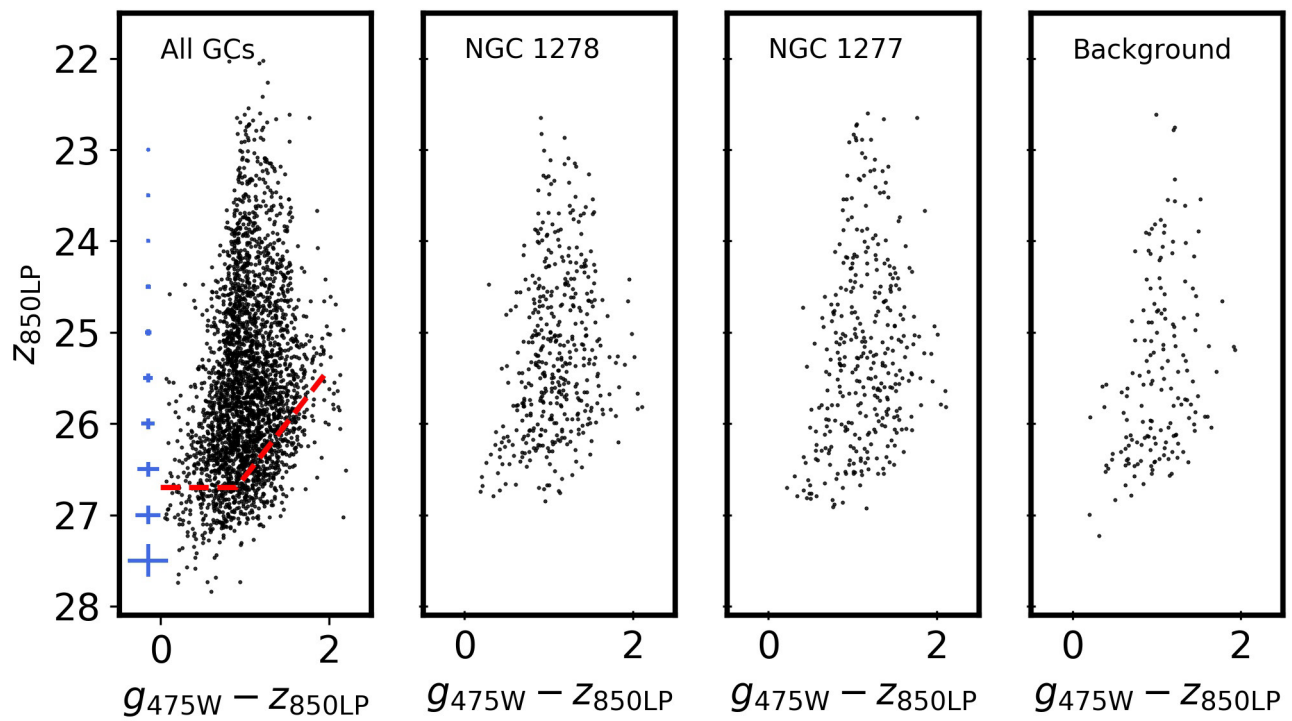
29. Bertin, E. & Arnout, S. SExtractor: software for source extraction. *Astron. Astrophys. Suppl. Ser.* **117**, 393–404 (1996).
30. Bertin, E. Automated morphometry with SExtractor and PSFEx. *ASP Conf. Ser.* **442**, 435–438 (2011).
31. Schlafly, E. F. & Finkbeiner, D. P. Measuring reddening with Sloan Digital Sky Survey stellar spectra and recalibrating SFD. *Astrophys. J.* **737**, 103 (2011).
32. Bohlin, R. C. Perfecting the photometric calibration of the ACS CCD cameras. *Astron. J.* **152**, 60 (2016).
33. Scott, D. W. *Multivariate Density Estimation: Theory, Practice, and Visualization* 7–11 (John Wiley & Sons, 1992).
34. Pota, V. et al. The SLUGGS survey: kinematics for over 2500 globular clusters in 12 early-type galaxies. *Mon. Not. R. Astron. Soc.* **428**, 389–420 (2013).
35. Sérsic, J. L. The influence of the atmospheric and instrumental dispersion on the brightness distribution in a galaxy. *Bol. Asoc. Argentina Astron.* **6**, 41–43 (1963).
36. Ferrarese, L. et al. A database of Cepheid distance moduli and tip of the red giant branch, globular cluster luminosity function, planetary nebula luminosity function, and surface brightness fluctuation data useful for distance determinations. *Astrophys. J. Suppl. Ser.* **128**, 431–459 (2000).
37. Secker, J. & Harris, W. E. A maximum likelihood analysis of globular cluster luminosity distributions in the Virgo ellipticals. *Astron. J.* **105**, 1358–1368 (1993).
38. Jordán, A. et al. The ACS Virgo Cluster Survey. XII. The luminosity function of globular clusters in early-type galaxies. *Astrophys. J. Suppl. Ser.* **171**, 101–145 (2007).

39. Peng, E. W. *et al.* The ACS Virgo Cluster Survey. XV. The formation efficiencies of globular clusters in early-type galaxies: the effects of mass and environment. *Astrophys. J.* **681**, 197–224 (2008).
40. Harris, W. E. *et al.* Globular cluster systems in brightest cluster galaxies. bimodal metallicity distributions and the nature of the high-luminosity clusters. *Astrophys. J.* **636**, 90–114 (2006).
41. Strader, J., Brodie, J. P., Spitler, L. & Beasley, M. A. Globular clusters in Virgo ellipticals: unexpected results for giants and dwarfs from advanced camera for surveys imaging. *Astron. J.* **132**, 2333–2345 (2006).
42. Mieske, S. *et al.* The ACS Fornax Cluster Survey. IX. The color–magnitude relation of globular cluster systems. *Astrophys. J.* **710**, 1672–1682 (2010).
43. Poggianti, B. M. *et al.* Superdense galaxies and the mass-size relation at low redshift. *Astrophys. J.* **762**, 77 (2013).
44. Valentinuzzi, T. *et al.* Superdense massive galaxies in wings local clusters. *Astrophys. J.* **712**, 226–237 (2010).
45. Peralta de Arriba, L., Quilis, V., Trujillo, I., Cebrián, M. & Balcells, M. Massive relic galaxies prefer dense environments. *Mon. Not. R. Astron. Soc.* **461**, 156–163 (2016).
46. Damjanov, I., Zahid, H. J., Geller, M. J. & Hwang, H. S. The environment of massive quiescent compact galaxies at $0.1 < z < 0.4$ in the COSMOS field. *Astrophys. J.* **815**, 104 (2015).
47. Stringer, M., Trujillo, I., Dalla Vecchia, C. & Martinez-Valpuesta, I. A cosmological context for compact massive galaxies. *Mon. Not. R. Astron. Soc.* **449**, 2396–2404 (2015).
48. Trujillo, I. *et al.* Superdense Massive Galaxies in the Nearby Universe. *Astrophys. J.* **692**, L118–L122 (2009).
49. Taylor, E. N. *et al.* On the dearth of compact, massive, red sequence galaxies in the local Universe. *Astrophys. J.* **720**, 723–741 (2010).
50. Leauthaud, A. *et al.* New constraints on the evolution of the stellar- to-dark matter connection: a combined analysis of galaxy-galaxy lensing, clustering, and stellar mass functions from $z=0.2$ to $z=1$. *Astrophys. J.* **744**, 159 (2012).
51. Kirby, E. N. *et al.* The Universal stellar mass-stellar metallicity relation for dwarf galaxies. *Astrophys. J.* **779**, 102 (2013).
52. Georgiev, I. Y., Puzia, T. H., Goudfrooij, P. & Hilker, M. Globular cluster systems in nearby dwarf galaxies – III. Formation efficiencies of old globular clusters. *Mon. Not. R. Astron. Soc.* **406**, 1967–1984 (2010).



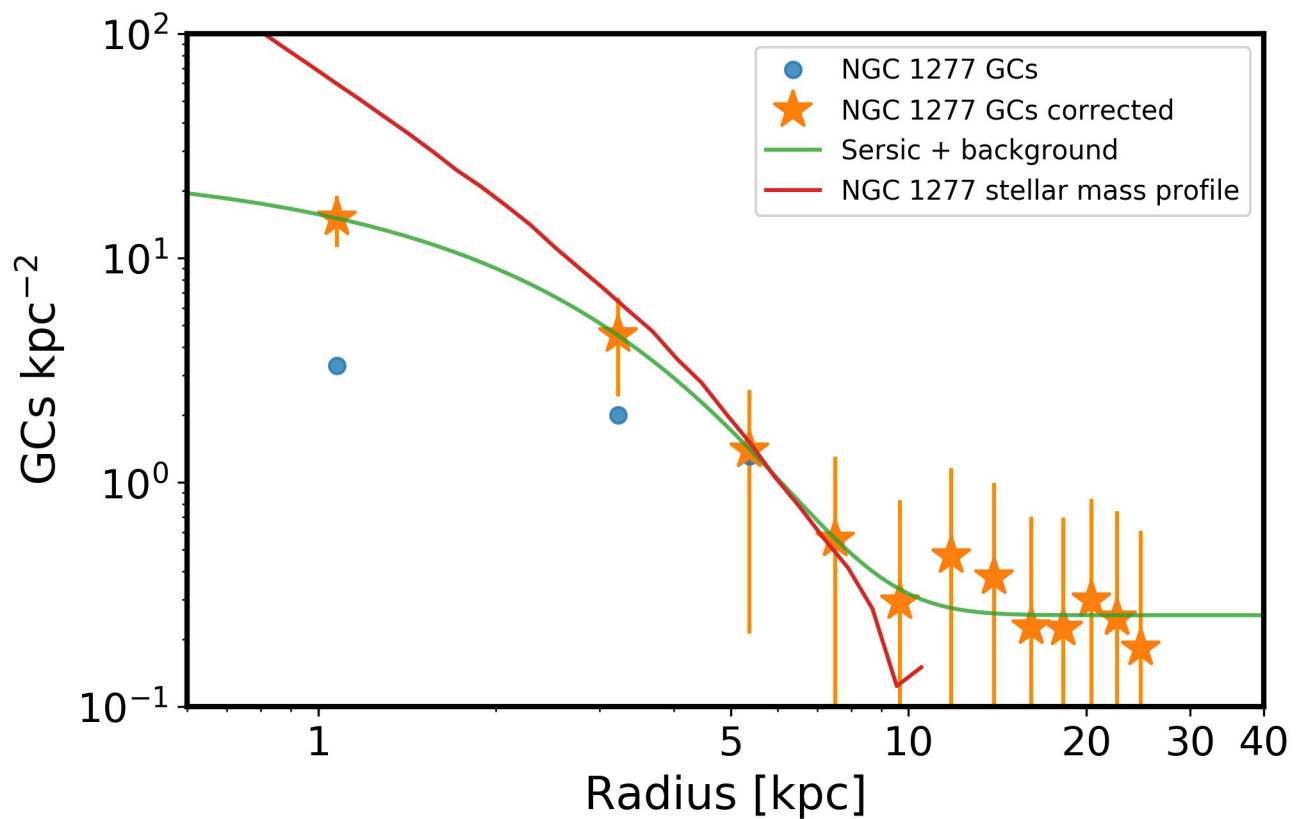
Extended Data Figure 1 | Colour-composite HST image of the massive relic galaxy NGC 1277. The image was constructed from $g_{475W} + r_{625W} + z_{850LP}$ bands. The g_{475W} and z_{850LP} imaging was obtained with the HST programme GO-14215 and the r_{625W} imaging with the programme GO-10546. NGC 1277 is the best example found so far in the nearby Universe with characteristics equivalent to those of the first massive galaxies to form more than 11 Gyr ago. The image

is oriented with north pointing up and east to the left. The field of view is $42.2 \text{ arcsec} \times 30.6 \text{ arcsec}$, corresponding to a physical scale of $14.5 \text{ kpc} \times 10.5 \text{ kpc}$ at our adopted distance of 73.3 Mpc to the galaxy. The images have been scaled with logarithmic intensity to highlight the various structures in the galaxy. The vast majority of point sources surrounding NGC 1277 are clusters associated with the galaxy. Two example clusters have been marked with ticks.



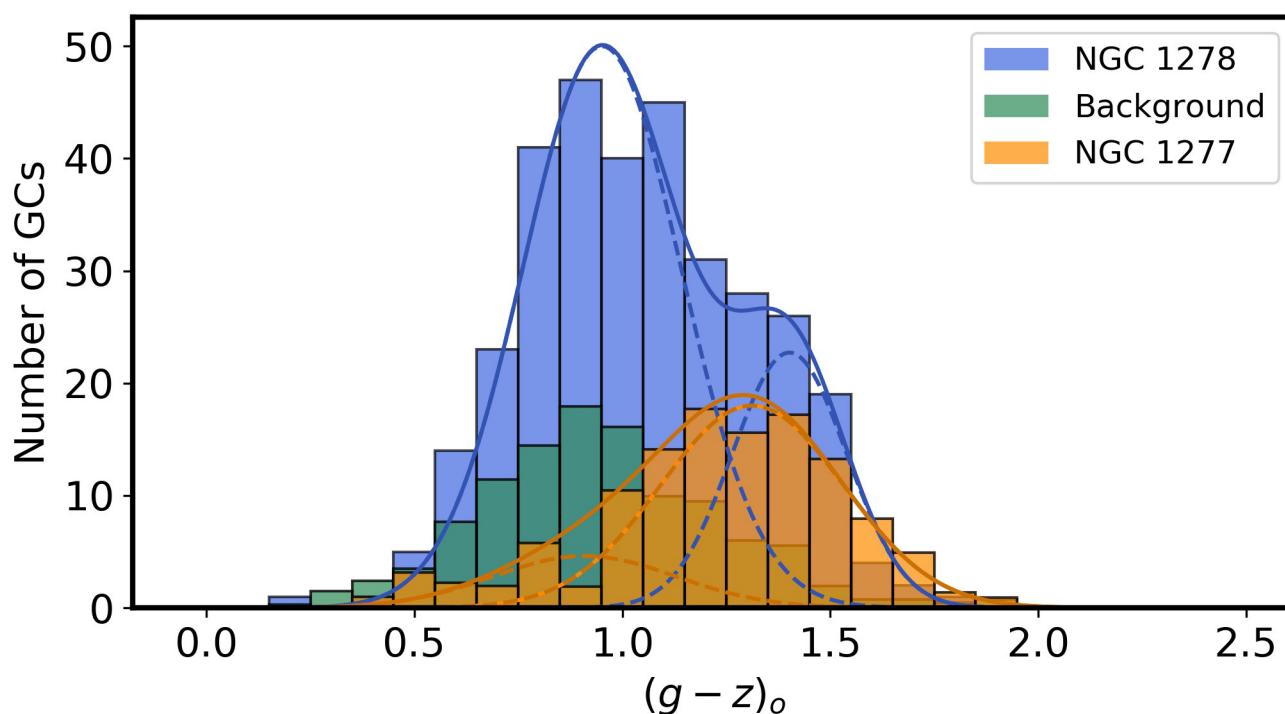
Extended Data Figure 2 | Colour-magnitude diagrams of cluster candidates. From left to right, the plots show all point sources in the field, NGC 1278 clusters, raw (not background-corrected) NGC 1277

clusters and an example background field. Also shown are the photometric uncertainties (1σ ; blue crosses) based on our artificial point-source test, and the 100% completeness limit (red dashed lines; see Methods).



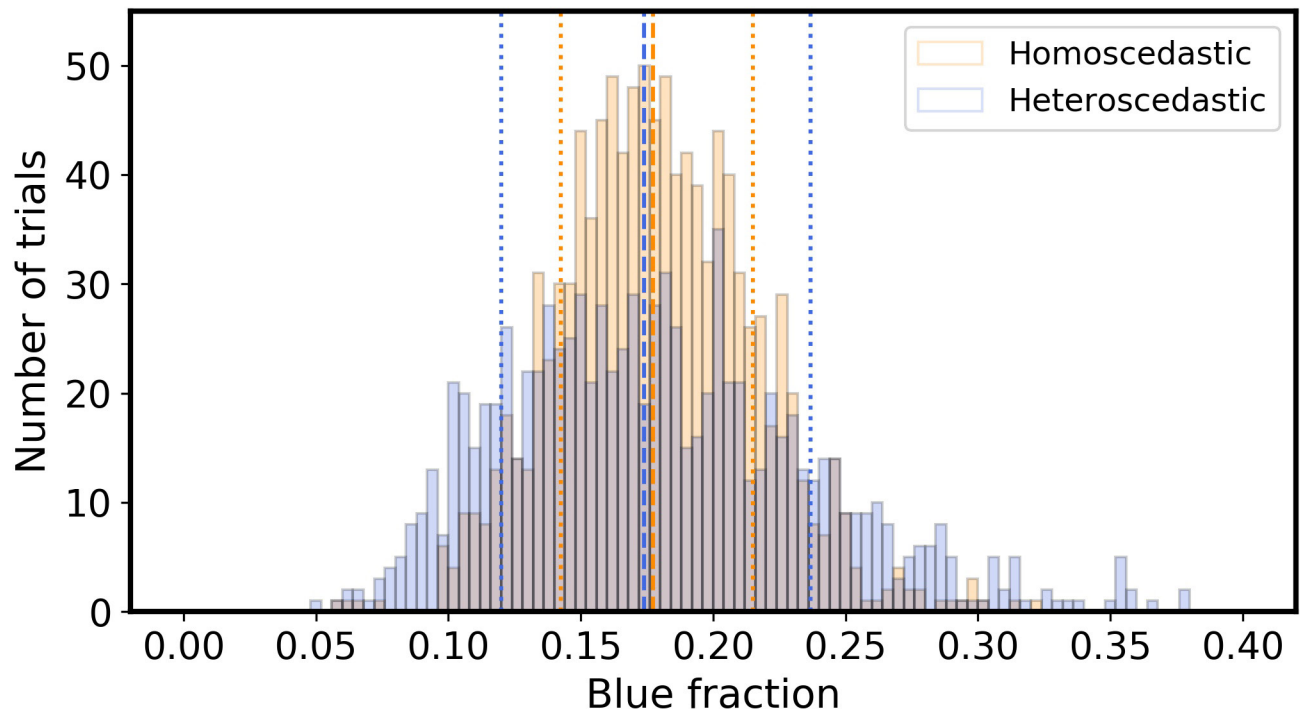
Extended Data Figure 3 | Surface-density profile of NGC 1277 clusters. The figure shows the raw (blue circles) and completeness-corrected (orange stars) counts. Uncertainties are 16th–84th percentiles. Within the range 3–10 kpc, the clusters closely follow the light distribution of the galaxy. This is typical of red clusters. The background is determined from

the Sérsic function fits (which include a background term; green line) to the cluster data and defines the radial extent of the cluster system, which ends at 11 kpc (about $10R_c$). The stellar-mass density profile¹⁵ (red line) has arbitrary normalization. GCs, globular clusters.

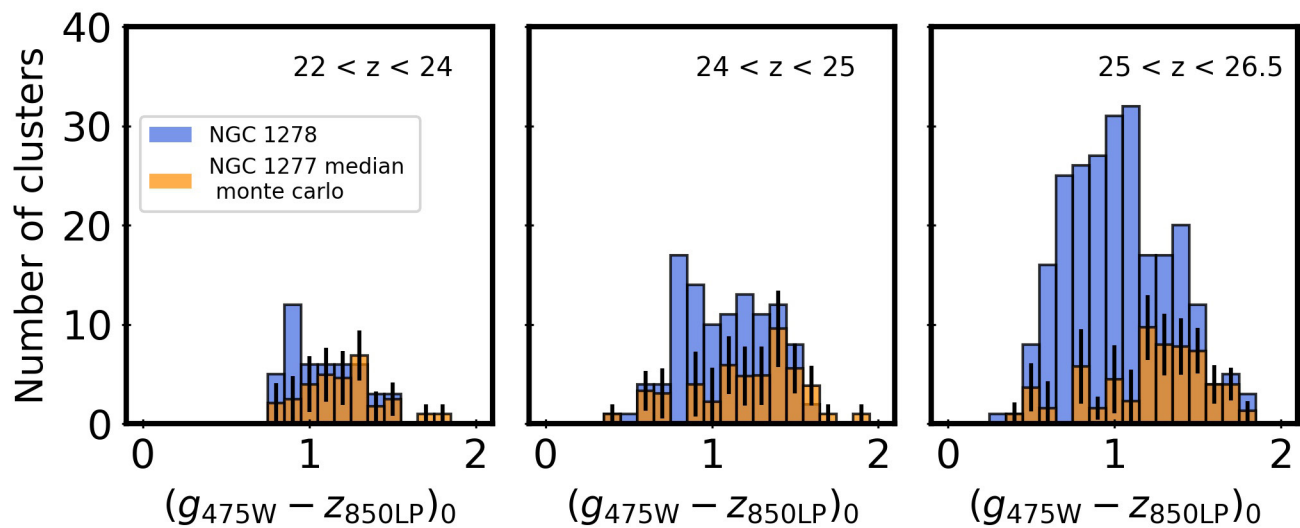


Extended Data Figure 4 | Colour distribution of clusters in NGC 1277 compared with that for the companion galaxy in projection NGC 1278. The figure shows that NGC 1278 (blue) has a strong peak of blue clusters that is not seen in NGC 1277 (orange). The expected background contamination of NGC 1278 clusters and of intracluster clusters to the cluster system of NGC 1277, with which we have corrected the colour distribution of NGC 1277, is also shown (green). The predominantly blue

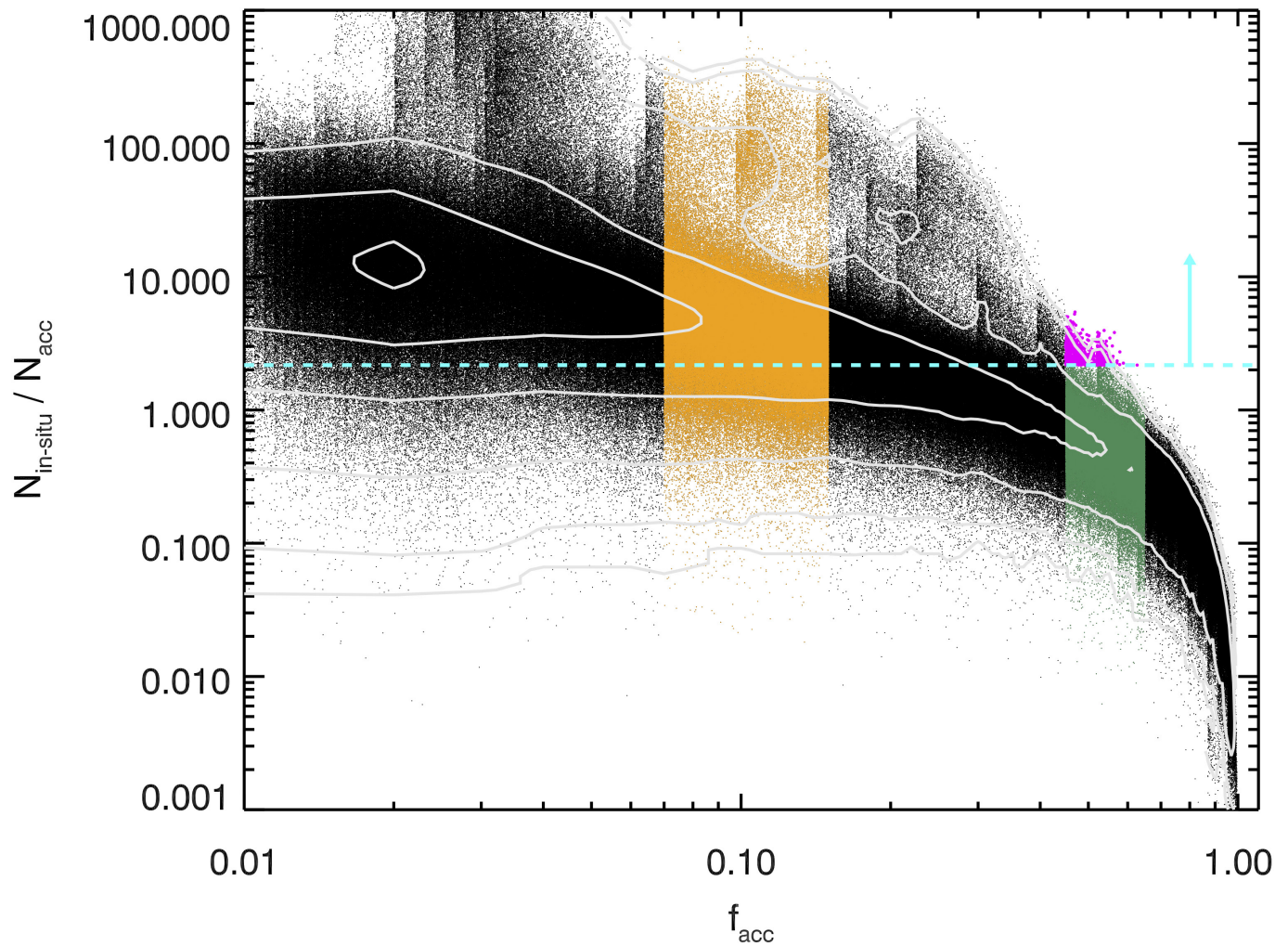
colours of the contaminating clusters is expected because they correspond to the outskirts of NGC 1278 (see Fig. 1). Solid curves are two-Gaussian solutions from the GMM code²⁸; dashed curves are individual Gaussian components from GMM. A single Gaussian fit to the NGC 1277 clusters with GMM gives $\langle g_{475W} - z_{850LP} \rangle = 1.22 \pm 0.03$, with a FWHM of 0.28 ± 0.02 .



Extended Data Figure 5 | Distribution of the blue cluster fraction from Monte Carlo simulations. The vertical dashed lines indicate the medians of the distributions and the vertical dotted lines show the 16th and 84th percentiles.

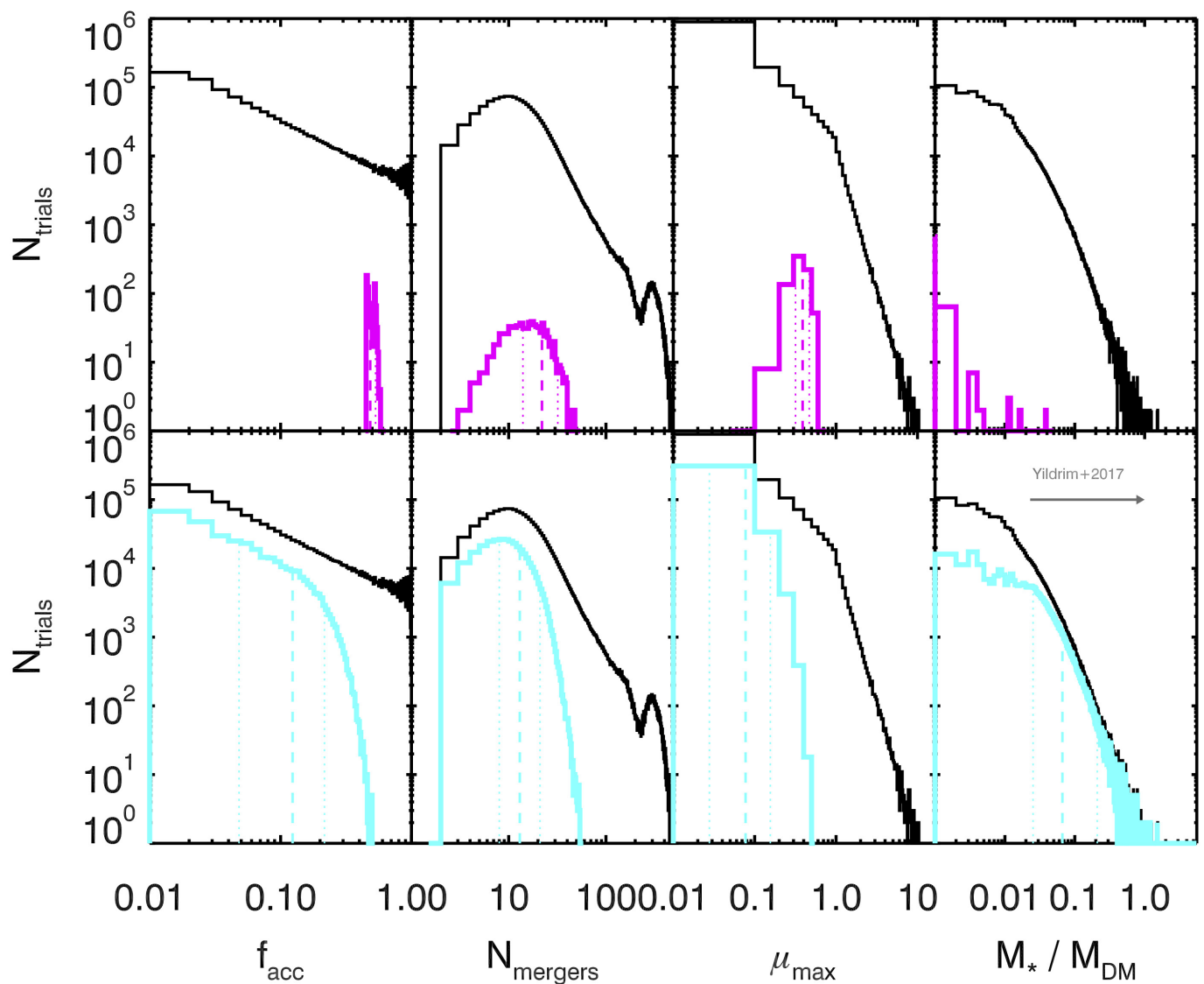


Extended Data Figure 6 | Colour distributions of NGC 1277 and NGC 1278 in three magnitude bins. The bin values for the NGC 1277 clusters represent the medians of the bin values from Monte Carlo simulations. Uncertainties are the 16th–84th percentiles of the distributions.



Extended Data Figure 7 | Accretion models for the build-up of the NGC 1277 cluster system. Each point is a single merger history for NGC 1277, characterized by the total fraction of accreted stellar mass versus the number ratio of *in situ* to accreted clusters for that realization. The observed ratio for NGC 1277 is indicated by the cyan dashed line ($N_{\text{blue}}/N_{\text{red}} = 0.21$) and is characteristic of galaxies with merger histories leading to $f_{\text{acc}} \approx 10\%$ (orange). The cyan arrow indicates that the

in situ fraction is a lower limit. Galaxies with comparable stellar mass to NGC 1277 in the ACS VCS tend to have equal numbers of blue and red clusters, which is more common for accretion histories of $f_{\text{acc}} \approx 50\%–90\%$ (green). The extremely rare (about 0.02%) high-mass-accretion realizations, which produce the observed blue/red fraction in NGC 1277, are shown in magenta. Contours represent 10%, 25%, 50%, 75%, 90% and 99% of the maximum of the two-dimensional distribution.



Extended Data Figure 8 | Predictions of galaxy properties from accretion models. From left to right, the black lines show the accretion fraction, the number of mergers, the maximum mass ratio of the merger and the final ratio of stellar to dark-matter mass for all of the merger realizations. The magenta distributions (top) show the high- f_{acc} merger histories that satisfy the constraint on the colour ratio of the clusters (which happens rarely; see, for example, Extended Data Fig. 7). The cyan

distributions (bottom) show the low- f_{acc} merger histories that satisfy the observed colour ratio of the clusters and the S_N constraints. The high-accretion-fraction models (magenta) also predict a dark-matter-halo mass for NGC 1277 that is larger than the observational constraints¹⁷ (indicated by the arrow in the bottom right panel). The vertical dashed lines represent the median of the distributions and the vertical dotted lines show the 16th and 84th percentiles.

Image reconstruction by domain–transform manifold learning

Bo Zhu^{1,2,3}, Jeremiah Z. Liu⁴, Stephen F. Cauley^{1,2}, Bruce R. Rosen^{1,2} & Matthew S. Rosen^{1,2,3}

Image reconstruction is essential for imaging applications across the physical and life sciences, including optical and radar systems, magnetic resonance imaging, X-ray computed tomography, positron emission tomography, ultrasound imaging and radio astronomy^{1–3}. During image acquisition, the sensor encodes an intermediate representation of an object in the sensor domain, which is subsequently reconstructed into an image by an inversion of the encoding function. Image reconstruction is challenging because analytic knowledge of the exact inverse transform may not exist *a priori*, especially in the presence of sensor non-idealities and noise. Thus, the standard reconstruction approach involves approximating the inverse function with multiple *ad hoc* stages in a signal processing chain^{4,5}, the composition of which depends on the details of each acquisition strategy, and often requires expert parameter tuning to optimize reconstruction performance. Here we present a unified framework for image reconstruction—automated transform by manifold approximation (AUTOMAP)—which recasts image reconstruction as a data-driven supervised learning task that allows a mapping between the sensor and the image domain to emerge from an appropriate corpus of training data. We implement AUTOMAP with a deep neural network and exhibit its flexibility in learning reconstruction transforms for various magnetic resonance imaging acquisition strategies, using the same network architecture and hyperparameters. We further demonstrate that manifold learning during training results in sparse representations of domain transforms along low-dimensional data manifolds, and observe superior immunity to noise and a reduction in reconstruction artefacts compared with conventional handcrafted reconstruction methods. In addition to improving the reconstruction performance of existing acquisition methodologies, we anticipate that AUTOMAP and other learned reconstruction approaches will accelerate the development of new acquisition strategies across imaging modalities.

The paradigm shift from manual to automatic feature extraction in a host of machine learning tasks including speech recognition⁶ and image classification⁷ has demonstrated the advantage of allowing real-world data to guide efficient representation through a structured training process. This strategy is mirrored in biological organisms for refining visual perception in a process known as perceptual learning⁸. Human visual reconstruction of time-domain neural codes into the percept image is trained through experience during cognitive development into adulthood. This conditioning on prior data has been shown to be critical to robust performance in low signal-to-noise settings⁹, which are fundamentally challenging for artificial imaging systems across disciplines and applications. In contemporary medical imaging, faithful reconstruction of noisy image acquisitions is of particular importance as the clinical push for faster scanning increasingly relies on acquisition strategies that result in a reduction of the signal-to-noise ratio, be they undersampled magnetic resonance imaging (MRI), or low-dose X-ray computed tomography imaging.

Inspired by the perceptual learning archetype, we describe here a data-driven unified image reconstruction approach, which we call AUTOMAP, that learns a reconstruction mapping between the sensor-domain data and image-domain output (Fig. 1a). As this mapping is trained, a low-dimensional joint manifold of the data in both domains is implicitly learned (Fig. 1b), capturing a highly expressive representation that is robust to noise and other input perturbations.

We implemented the AUTOMAP unified reconstruction framework with a deep neural network feed-forward architecture composed of fully connected layers followed by a sparse convolutional autoencoder (Fig. 1c). The fully connected layers approximate the between-manifold projection from the sensor domain to the image domain. The convolutional layers extract high-level features from the data and force the image to be represented sparsely in the convolutional-feature space. Our network operates similarly to the denoising autoencoder described previously¹⁰, but rather than finding an efficient representation of the identity to map $f(x) = \phi_x \circ \phi_x^{-1}(x) = x$ over the manifold of inputs \mathcal{X} (where ϕ_x maps the intrinsic coordinate system of \mathcal{X} to Euclidean space near x), AUTOMAP determines both a between-manifold projection g from \mathcal{X} (the manifold of sensor inputs) to \mathcal{Y} (the manifold of output images), and a manifold mapping ϕ_y to project the image from manifold \mathcal{Y} back to the representation in Euclidean space. A composite inverse transformation $f(x) = \phi_y \circ g \circ \phi_x^{-1}(x)$ over the joint manifold $\mathcal{M}_{\mathcal{X}\mathcal{Y}} = \mathcal{X} \times \mathcal{Y}$ (Fig. 1b) is achieved. A full mathematical description of this manifold learning process is detailed in Methods.

In contrast to previous efforts that use neural networks to solve inverse functions^{11–13}, our approach searches for an inverse that best represents the data in a low-dimensional feature space determined by manifold learning as well as the trained sparse convolutional filters. Furthermore, AUTOMAP solves a generalized reconstruction problem and thus differs from work using neural networks to implement a specific image reconstruction task^{14–17}. These previous approaches use known properties of the canonical domain transform to formulate the neural network model, or perform the explicit transform before processing by a neural network used for image-space artefact reduction.

We demonstrate AUTOMAP image reconstruction using MRI as a model system, but we emphasize that our approach is applicable to image reconstruction problems across a broad range of modalities given the mathematical similarities of tomographic spatial encoding functions typically governed by Fredholm integral equations¹. The plethora of MRI acquisition strategies makes it a particularly appropriate platform to exhibit the flexibility of AUTOMAP reconstruction over a variety of encoding schemes. We first evaluated the performance of AUTOMAP alongside conventional methods in four nontrivial reconstruction tasks: (1) Radon projection imaging and model-based iterative reconstruction⁴; (2) spiral-trajectory k -space (rapid acquisition with non-Cartesian sampling) and conjugate-gradient sensitivity encoding (SENSE) reconstruction employing non-uniform fast Fourier transform (NUFFT) regridding⁵; (3) Poisson-disk undersampled k -space (incoherent sparse acquisition) and compressed sensing

¹A. A. Martinos Center for Biomedical Imaging, Massachusetts General Hospital, Boston, Massachusetts, USA. ²Harvard Medical School, Boston, Massachusetts, USA. ³Department of Physics, Harvard University, Cambridge, Massachusetts, USA. ⁴Department of Biostatistics, Harvard University, Cambridge, Massachusetts, USA.

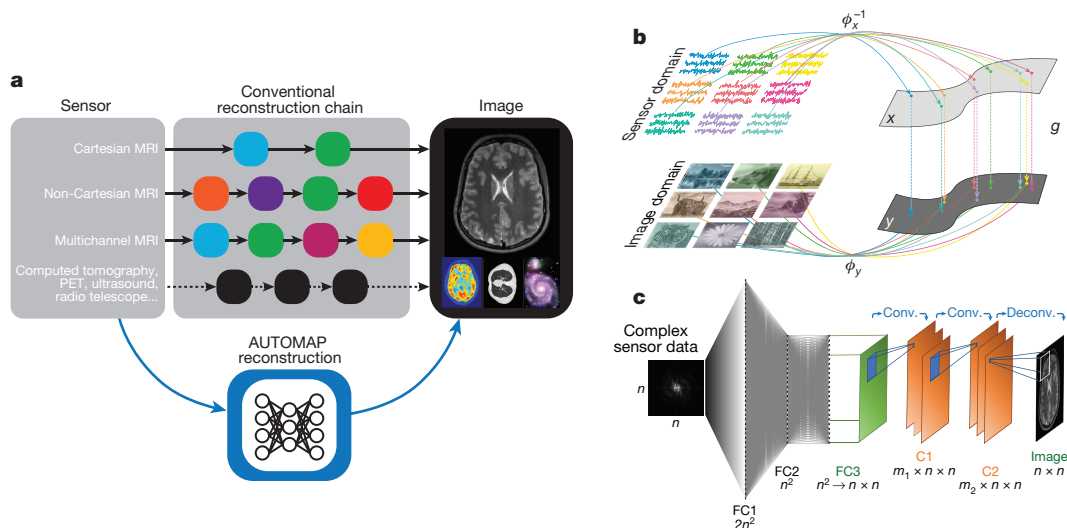


Figure 1 | Schematic representations of AUTOMAP image reconstruction.

a, Conventional image reconstruction is implemented with sequential modular reconstruction chains composed of handcrafted signal processing stages that may include discrete transforms (for example, Fourier, Hilbert or Radon), data interpolation techniques, nonlinear optimization, and various filtering mechanisms. AUTOMAP replaces this approach with a unified image reconstruction framework that learns the reconstruction relationship between sensor and image domain without expert knowledge. **b**, A mapping between sensor domain and image domain is determined via supervised learning of sensor (top) and image

reconstruction with wavelet sparsifying transform¹⁸; and (4) misaligned k -space (a commonplace sampling inaccuracy due to hardware limitations or physiologic effects) and the conventional inverse fast Fourier transform. Evaluation of the AUTOMAP network was performed on brain magnetic resonance images selected from the Human Connectome Project (HCP)¹⁹, which were transformed to the sensor domain according to the four encoding schemes (see Methods for data preparation details) and with varying levels of additive white Gaussian noise introduced so that we could observe reconstruction performance in noisy conditions.

All reconstruction tasks employed the same network architecture and hyperparameters—only the training data differed at the network input and output. To demonstrate AUTOMAP's generalizability in training dataset scope, all reconstruction tasks except the undersampled encoding were trained from datasets derived entirely from photographs of natural scenes from ImageNet²⁰ as schematically portrayed in Fig. 1b; for these acquisitions, the network was not exposed to any MRI or other medical images until the test phase (see Methods for data preparation and training details).

The results shown in Fig. 2 demonstrate the ability of AUTOMAP to reconstruct sensor-domain data across varying encoding acquisition strategies. We emphasize here that the reconstruction transforms emerged strictly from training on data samples, without higher-level knowledge (for example, mathematical transforms or domain representations) introduced at any stage. To learn a new reconstruction for a particular encoding acquisition, one simply needs to generate a training dataset with the encoding forward model. The ability of AUTOMAP to represent a variety of sophisticated transform functions with a single network architecture is grounded in the inherent universal approximation properties of nonlinear multilayer perceptron systems²¹.

Furthermore, AUTOMAP reconstructions exhibit superior noise immunity compared to those from conventional methods, as quantified by image signal-to-noise ratio and root-mean-squared error (RMSE) metrics (Fig. 2). Visual inspection of reconstructed images and error maps in Fig. 2 reveals that noise and reconstruction artefacts are diminished in AUTOMAP reconstructions compared to conventional reconstructions: streaking artefacts and white noise amplification for iterative

(bottom) domain pairs. The training process implicitly learns a low-dimensional joint manifold $\mathcal{X} \times \mathcal{Y}$ over which the reconstruction function $f(x) = \phi_y \circ g \circ \phi_x^{-1}(x)$ is conditioned. **c**, AUTOMAP is implemented with a deep neural network architecture composed of fully connected layers (FC1 to FC3) with hyperbolic tangent activations, followed by convolutional layers with rectifier nonlinearity activations that form a convolutional autoencoder. Our network contains m_1 and m_2 convolutional feature maps at C1 and C2 respectively. The convolution and deconvolution operations are labelled 'conv.' and 'deconv.', respectively. The dimensionality of the input to the network is $n \times n$. See Methods for model architecture details.

inverse-Radon²², noise amplification due to iterative reconstruction with NUFFT regridding of noisy samples²³, structured artefacts from noisy undersampled compressed sensing reconstruction²⁴, and Nyquist N/2 ghosting from misaligned sampling trajectories²⁵. Additive Gaussian noise was not injected during training; the noise immunity we observe was not trained explicitly, or imposed by predictive noise modelling, but rather emerged as a result of the manifold learning process extracting robust features of the data, leading to improvement in signal-to-noise ratio during reconstruction. This emphasis on modelling features of the signal rather than the noise characteristics to achieve high performance in low-signal-to-noise-ratio regimes is consistent with the neural mechanisms underlying human visual perceptual learning²⁶.

We next examined the hidden-layer activity of our AUTOMAP network during the feed-forward reconstruction process. We trained AUTOMAP using training data derived from either ImageNet, HCP brain images, or random-valued Gaussian noise without any real-world image structure. Each trained network was then used to reconstruct the fully sampled Cartesian k -space of a single brain image (Extended Data Fig. 2). The activation values of the hidden-layer FC2 (Fig. 1c) are plotted in Fig. 3a–c. As the training moves from general (Fig. 3a) to specific (Fig. 3c), we observe the hidden-layer activity exhibiting greater sparsity, indicating successful extraction of robust features²⁷, consistent with the noise immunity observed in our experiments. We note that fully connected hidden-layer sparsity was not explicitly imposed (that is, not enforced by a penalty in the loss function), but emerged naturally through the training process. A normalized histogram of the hidden-layer activations is shown in Fig. 3d. A representative set of the convolutional kernels applied to feature maps in layer C2 (Fig. 1c) is shown in Fig. 3h. Processing by the convolutional layers is similar to that of compressed sensing, except that instead of assuming an explicit sparsifying transform (for example, wavelet), AUTOMAP simultaneously learns a sparse convolutional domain and its sparse representations through a joint optimization (see Methods for details).

We then studied the weight parameters of each trained network using a t -distributed stochastic neighbour embedding (t-SNE) analysis²⁸ (Fig. 3e–g), which embeds a high-dimensional dataset into a low-dimensional space for visualization. Here we visualize the spatial

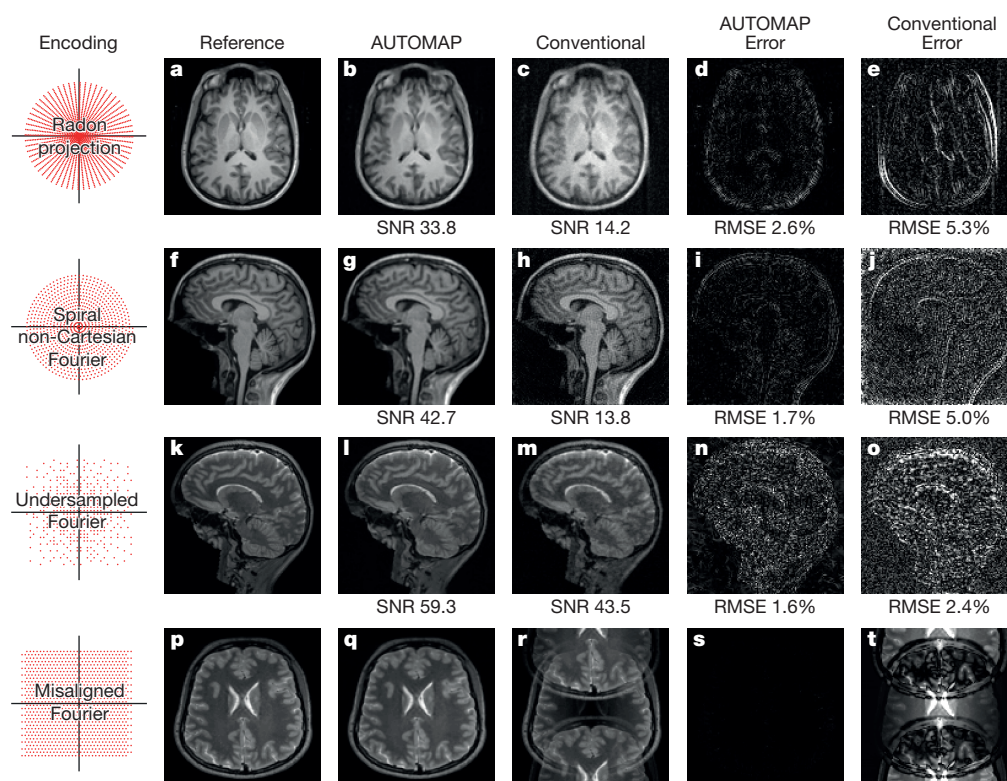


Figure 2 | Reconstruction performance of AUTOMAP compared with conventional techniques. Reference brain images were encoded into sensor-domain sampling strategies with varying levels of additive white Gaussian noise and reconstructed with both AUTOMAP and conventional approaches. **a–e**, Radon projection encoding compared with model-based iterative reconstruction. **f–j**, Spiral k -space encoding compared with conjugate-gradient SENSE reconstruction with NUFFT regridding. **k–o**, Poisson-disk undersampled (40%) Cartesian k -space encoding

compared with compressed sensing reconstruction using the wavelet sparsifying transform. **p–t**, Mismatched Cartesian k -space, compared with conventional inverse fast Fourier transform. Image magnitude signal-to-noise ratios (SNRs) and error maps (with root-mean-squared error (RMSE) calculations) with respect to reference ground truth images are also shown. For each encoding experiment, both error maps are windowed to the same level. The same network architecture was used for all AUTOMAP reconstructions.

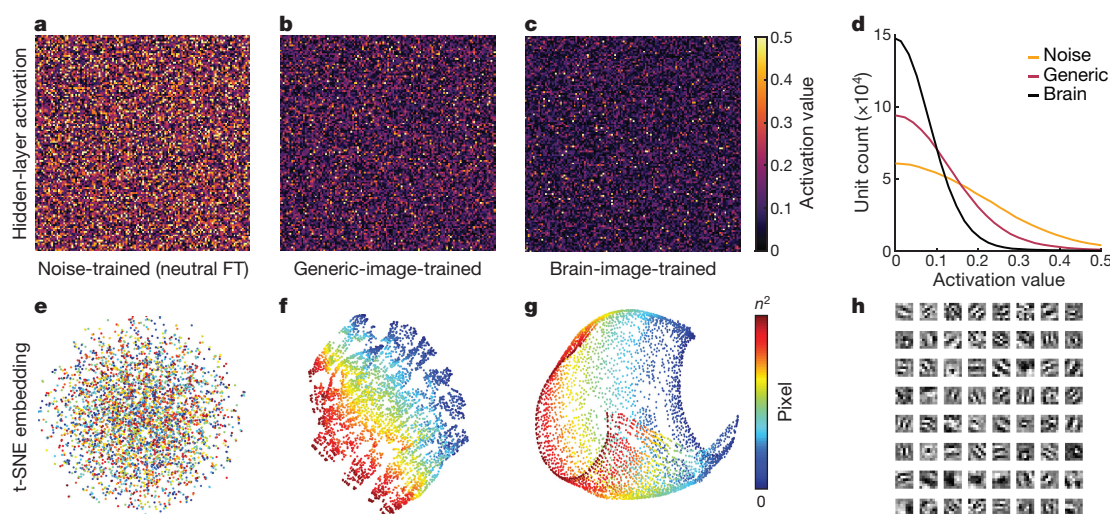


Figure 3 | Analysis of AUTOMAP neural networks. AUTOMAP was trained on three separate datasets for a Cartesian k -space encoding: generic natural images, brain images, and random-valued noise without any real-world image structure (see Methods for training details). **a–c**, Activation values of the fully connected hidden layer (FC2 in Fig. 1c) for each trained network while reconstructing the same k -space of a brain image. The noise-trained network generates high-amplitude and widely distributed hidden layer activation values (**a**), while the networks trained on generic images (**b**) and brain images (**c**) exhibit greater sparsity, indicating efficient processing of input data due to successful feature extraction when trained on relevant data. **d**, Histogram of FC2 activation

values for the three networks, accumulated over 100 brain-image k -space reconstructions. **e–g**, Three-dimensional t-SNE embedding of network weights from FC2 to FC3 for the differently trained networks (see Methods for t-SNE analysis details). The t-SNE of the noise-trained network, agnostic to real-world image structure, exhibits disorganized structure (**e**), in contrast to **f** and **g**, which reflect the local spatial correlation that exists in real-world images. The domain-specific training of the brain-trained network shows the highest similarity between weights to two-dimensional neighbours for all pixel locations (**g**). **h**, Representative sparse convolutional kernels of the final convolutional stage (C2–Image) learned from training on brain images.

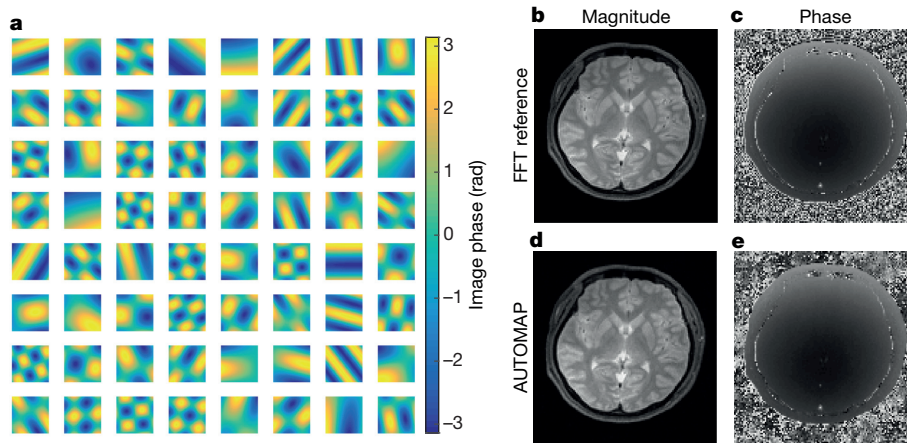


Figure 4 | Learning reconstruction of phase for *in vivo* data. The inclusion of synthetic phase to the training dataset enables AUTOMAP to properly reconstruct both the magnitude and phase. **a**, Synthetic k -space data for training was generated from HCP magnitude images by phase

modulation with two-dimensional sinusoids of varying spatial frequencies. **b–e**, After training, the magnitude (**b**) and phase (**c**) of T2-weighted raw k -space acquired from a human subject are properly reconstructed by AUTOMAP (**d**, **e**). (FFT, fast Fourier transform.)

relationship of trained network weights, particularly from FC2 to the pre-convolutional FC3 image layer (see Methods for details). Figure 3d shows that the t-SNE embedding of the noise-trained network weights is highly unorganized and unbiased with respect to pixel location; this is unsurprising because the network learns a ‘pure’ or ‘neutral’ Fourier transform that does not recognize the local spatial correlation that exists in real-world images²⁹. The generic-image-trained model captures this local spatial correlation, so the weights of neighbouring pixels are more similar than the weights of pixels further apart, as shown by the t-SNE embedding (Fig. 3f). This feature is most clearly exhibited by the t-SNE of the brain-trained network weights (Fig. 3g), which are organized into a two-dimensional sheet within the three-dimensional embedding space, demonstrating extremely high similarity between the weights of two-dimensional neighbours for all pixel locations.

We then demonstrated AUTOMAP’s ability to learn reconstruction of image phase from complex-valued sensor data by including phase-modulated data in the network training. A phase-modulated training set was created by generating synthetic phase patterns (examples shown in Fig. 4a) to modulate the magnitude-only training images collected from the HCP database before being encoded in k -space. Using the same k -space data as input, we trained separate AUTOMAP networks to reconstruct magnitude and phase with their respective target training images, and validated this by reconstructing *in vivo* k -space raw data taken from a human subject on a 3-tesla (3T) MRI scanner (Fig. 4d, e); see Methods for acquisition details. This reconstruction can also be performed on one larger network with concatenated magnitude and phase data. This phase-modulation training allows public medical image databases and private PACS (picture archiving and communication system) repositories to be used for training, despite the typical absence of phase data in these datasets. Furthermore, our results show that parameterized influences on the input signal can be simulated for training and subsequently disentangled by AUTOMAP, which may be useful in sophisticated reconstruction problems such as automated motion compensation (see Methods and Extended Data Fig. 4 for more detailed discussion).

A potential concern in the reconstruction of real-world experimentally acquired sensor data is whether the AUTOMAP network would overfit to the ideal sampling parameters used during training and as a result be overly sensitive to sampling deviation during actual acquisitions. We quantified the effect of divergence from the nominal sampling trajectory with Monte Carlo simulations of varying amplitudes of trajectory error (Fig. 5a) from a spiral acquisition, and measure the resulting reconstruction RMSE from the ground truth reference. To examine a broad range of potential errors with realistic trajectories, we measured the actual trajectory during a spiral acquisition and

computed the difference vectors of samples between the actual and ideal designed trajectories. This was used to scale the vector magnitudes to generate offset sampling trajectories. We tested errors from zero deviation (perfect match) to four times the measured deviation, and found that AUTOMAP’s reconstruction error smoothly increased as a function of trajectory error, similar to a conventional NUFFT reconstruction’s error curve (Fig. 5b), demonstrating reasonable robustness to trajectory deviation. Although the AUTOMAP error curve was slightly steeper, AUTOMAP still achieved better reconstruction accuracy than did NUFFT, out to very large trajectory errors, more than 3.5 times larger than the measured experimental deviation (Fig. 5b) from a commercial 3T MRI scanner.

These simulation results are consistent with the reconstruction performance on real scanner data acquired from human subjects. Figure 5c, d shows AUTOMAP and NUFFT reconstructions of a 10-interleave spiral magnetic resonance acquisition, in which both methods assume the nominal trajectory that deviates from the actual experimental scan trajectory. Although there is no ground truth with which to calculate reconstruction error, image signal-to-noise ratio was measured to be higher in the AUTOMAP output (21.6 versus 17.6). Figure 5e, f displays windowed versions of Fig. 5c, d, revealing coherent object-dependent and ringing artefacts in the NUFFT reconstruction (Fig. 5f); these are much reduced in the AUTOMAP reconstruction, primarily exhibiting standard Gaussian white noise (Fig. 5e).

Finally, we demonstrate reconstruction of multichannel magnetic resonance data acquired on a clinical 3T scanner with 15.5 times undersampling (acceleration factor $R = 4 \times 4$ uniform with low frequency region) retrospective undersampling (Fig. 5g–k). In comparison to a conventional SENSE reconstruction (Fig. 5i), AUTOMAP (Fig. 5h) demonstrates reduced noise and reconstruction artefacts, which can clearly be observed in the error maps (Fig. 5j, k) and quantified by a reduction in RMSE from 10.8% to 6.72%). Further acquisition and reconstruction details can be found in Methods.

At its core, AUTOMAP is a conceptual approach for trained image reconstruction with manifold learning; the specific neural network implementation presented here is not the only possible implementation, but a first demonstration that can be extended and improved upon in many directions. In Methods we discuss the application of AUTOMAP to other reconstruction problems and ways to address practical implementation challenges. As an example of applicability beyond MRI, human ¹⁸F-fluorodeoxyglucose (FDG) positron emission tomography (PET) data are reconstructed with AUTOMAP in Extended Data Fig. 5.

AUTOMAP provides a new paradigm for image reconstruction that learns a reconstruction function for arbitrary acquisition strategies,

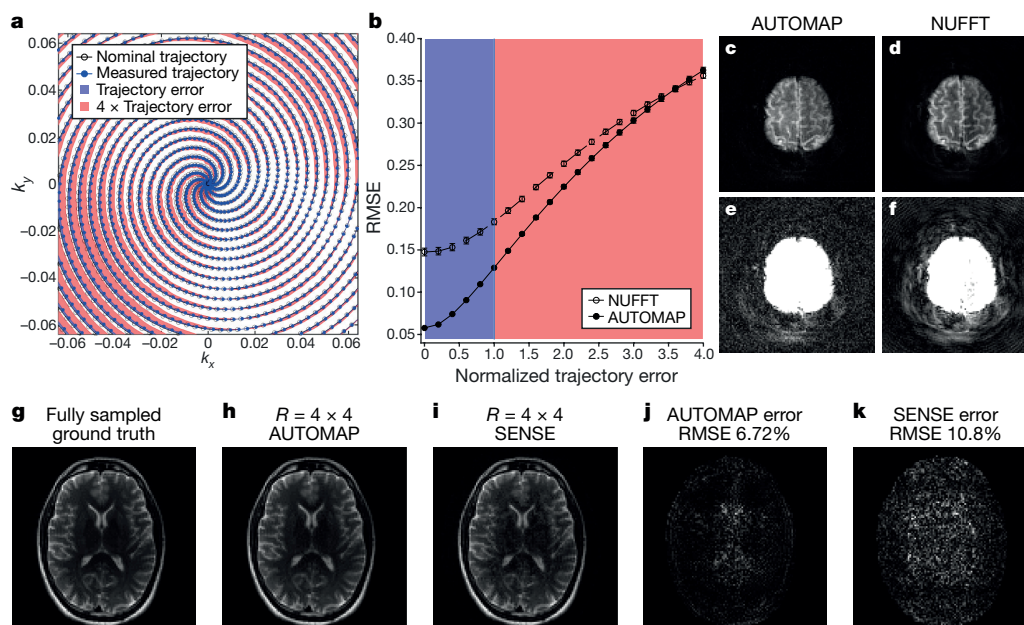


Figure 5 | Performance of AUTOMAP in real-world acquisitions. **a**, Plot of the family of k -space trajectory deviations used in Monte Carlo reconstruction analysis as described in the text. **b**, RMSE of AUTOMAP and NUFFT Monte Carlo reconstructions as a function of normalized trajectory error, with unity error corresponding to the measured experimental trajectory deviation in a 3T clinical scanner. Error bars indicate standard error of the mean. **c**, **d**, AUTOMAP and NUFFT reconstructions of a real-world 3T magnetic resonance spiral

acquisition; the mean image signal-to-noise ratio is improved in the AUTOMAP reconstruction (21.6 versus 17.6). **e**, **f**, The presence of reconstruction artefacts is also reduced, as revealed in the windowed images. **g**–**i**, AUTOMAP and SENSE reconstructions of a real-world $R=4 \times 4$ undersampled 32-channel 3T magnetic resonance acquisition are displayed and compared with the fully sampled reference. **j**, **k**, Error maps are windowed to the same intensity level. The RMSE reconstruction error is reduced with AUTOMAP compared to SENSE (6.72% versus 10.8%).

conditioned upon low-dimensional features of real-world data to improve artefact reduction and reconstruction accuracy for noisy and undersampled acquisitions. We anticipate that the noise robustness attainable with our approach will improve imaging quality and speed for a broad range of applications exhibiting low signal-to-noise ratio, including low-dose X-ray computed tomography³⁰, low-light charge-coupled devices³¹, large-baseline radio astronomy³² and rapid volumetric optical coherence tomography³³. Finally, we also anticipate that the AUTOMAP paradigm will enable the development of new classes of acquisition strategies across imaging modalities as the reconstruction of arbitrary encoding schemes can be learned without domain expert knowledge.

Online Content Methods, along with any additional Extended Data display items and Source Data, are available in the online version of the paper; references unique to these sections appear only in the online paper.

Received 12 April 2017; accepted 23 January 2018.

- Grangeat, P. *Tomography* (John Wiley & Sons, 2013).
- Gull, S. F. & Daniell, G. J. Image reconstruction from incomplete and noisy data. *Nature* **272**, 686–690 (1978).
- Zeng, G. L. *Medical Image Reconstruction* (Springer, 2010).
- Yu, Z., Thibault, J.-B., Bouman, C. A., Sauer, K. D. & Hsieh, J. Fast model-based X-ray CT reconstruction using spatially nonhomogeneous ICD optimization. *IEEE Trans. Image Process.* **20**, 161–175 (2011).
- Pruessmann, K. P., Weiger, M., Börner, P. & Boesiger, P. Advances in sensitivity encoding with arbitrary k -space trajectories. *Magn. Reson. Med.* **46**, 638–651 (2001).
- Hinton, G. et al. Deep neural networks for acoustic modeling in speech recognition: the shared views of four research groups. *IEEE Signal Process. Mag.* **29**, 82–97 (2012).
- Krizhevsky, A., Sutskever, I. & Hinton, G. E. ImageNet classification with deep convolutional neural networks. *Adv. Neural Inf. Process. Syst.* 1097–1105 (2012).
- Gilbert, C. D., Sigman, M. & Crist, R. E. The neural basis of perceptual learning. *Neuron* **31**, 681–697 (2001).
- Lu, Z.-L., Hua, T., Huang, C.-B., Zhou, Y. & Doshier, B. A. Visual perceptual learning. *Neurobiol. Learn. Mem.* **95**, 145–151 (2011).
- Vincent, P., Larochelle, H., Bengio, Y. & Manzagol, P.-A. Extracting and composing robust features with denoising autoencoders. In *Proc. 25th Int. Conf. on 'Machine Learning'* 1096–1103, <http://www.cs.toronto.edu/~larochel/publications/icml-2008-denoising-autoencoders.pdf> (2008).

- Ogawa, T., Kosugi, Y. & Kanada, H. Neural network based solution to inverse problems. In *IEEE World Congr. on 'Computational Intelligence'* Vol. 3, 2471–2476, <http://ieeexplore.ieee.org/document/687250/> (1998).
- Schiller, H. & Doerffer, R. Neural network for emulation of an inverse model operational derivation of Case II water properties from MERIS data. *Int. J. Remote Sens.* **20**, 1735–1746 (1999).
- Hoole, S. R. H. Artificial neural networks in the solution of inverse electromagnetic field problems. *IEEE Trans. Magn.* **29**, 1931–1934 (1993).
- Floyd, C. E. An artificial neural network for SPECT image reconstruction. *IEEE Trans. Med. Imaging* **10**, 485–487 (1991).
- Pelt, D. M. & Batenburg, K. J. Fast tomographic reconstruction from limited data using artificial neural networks. *IEEE Trans. Image Process.* **22**, 5238–5251 (2013).
- Jin, K. H., McCann, M. T., Froustey, E. & Unser, M. Deep convolutional neural network for inverse problems in imaging. *IEEE Trans. Image Process.* **26**, 4509–4522 (2017).
- Hammernik, K. et al. Learning a variational network for reconstruction of accelerated MRI data. *Magn. Reson. Med.* **79**, 3055–3071 (2017).
- Lustig, M., Donoho, D. & Pauly, J. M. Sparse MRI: the application of compressed sensing for rapid MR imaging. *Magn. Reson. Med.* **58**, 1182–1195 (2007).
- Fan, Q. et al. MGH-USC Human Connectome Project datasets with ultra-high b -value diffusion MRI. *Neuroimage* **124**, 1108–1114 (2016).
- Deng, J. et al. ImageNet: a large-scale hierarchical image database. In *IEEE Conf. on 'Computer Vision and Pattern Recognition'* 248–255, http://www.image-net.org/papers/imagenet_cvpr09.pdf (2009).
- Hornik, K., Stinchcombe, M. & White, H. Multilayer feedforward networks are universal approximators. *Neural Netw.* **2**, 359–366 (1989).
- Di Carli, M. F. & Lipton, M. J. *Cardiac PET and PET/CT Imaging* (Springer, 2007).
- Yang, Z. & Jacob, M. Mean square optimal NUFFT approximation for efficient non-Cartesian MRI reconstruction. *J. Magn. Reson.* **242**, 126–135 (2014).
- Virtue, P. & Lustig, M. On the empirical effect of Gaussian noise in under-sampled MRI reconstruction. Preprint at <https://arxiv.org/abs/1610.00410> (2016).
- Brown, R. W., Cheng, Y. C. N., Haacke, E. M., Thompson, M. R. & Venkatesan, R. *Magnetic Resonance Imaging: Physical Principles and Sequence Design* 2nd edn (Wiley, 2014).
- Gold, J., Bennett, P. J. & Sekuler, A. B. Signal but not noise changes with perceptual learning. *Nature* **402**, 176–178 (1999).
- Wright, J. et al. Sparse representation for computer vision and pattern recognition. *Proc. IEEE* **98**, 1031–1044 (2010).
- Maaten, L. V. D. & Hinton, G. Visualizing Data using t-SNE. *J. Mach. Learn. Res.* **9**, 2579–2605 (2008).
- Getis, A. in *Handbook of Applied Spatial Analysis* (eds Fisher, M. M. & Getis, A.) 255–278 (Springer, 2010).

30. Kubo, T. *et al.* Radiation dose reduction in chest CT: a review. *Am. J. Roentgenol.* **190**, 335–343 (2008).
31. Daigle, O., Djazovski, O., Laurin, D., Doyon, R. & Artigau, É. Characterization results of EMCCDs for extreme low-light imaging. In *Proc. SPIE on 'High Energy, Optical, and Infrared Detectors for Astronomy V'* Vol. 8453, 845303, <https://doi.org/10.1117/12.926385> (2012).
32. Girard, J. N. *et al.* Sparse representations and convex optimization as tools for LOFAR radio interferometric imaging. *J. Instrum.* **10**, C08013 (2015).
33. Lebed, E., Sarunic, M. V., Beg, M. F. & Mackenzie, P. J. Rapid volumetric OCT image acquisition using compressive sampling. *Opt. Exp.* **18**, 21003–21012 (2010).

Acknowledgements We acknowledge M. Michalski and the computational resources and assistance provided by the Massachusetts General Hospital (MGH) and the Brigham and Women's Hospital (BWH) Center for Clinical Data Science (CCDS). The CCDS is supported by MGH, BWH, the MGH Department of Radiology, the BWH Department of Radiology, and through industry partnership with NVIDIA. We also acknowledge the Center for Machine Learning at Martinos. We also thank J. Stockmann, J. Polimeni, D. E. J. Waddington and R. L. Walsworth for their comments on this manuscript, and B. Bilgic and C. Liao for their assistance in human subject data acquisition. We acknowledge C. Catana for providing raw PET data and for filtered back

projection and OSEM reconstructions. We also thank M. Haskell for providing the MRI motion encoding model. B.Z. was supported by the National Institutes of Health/National Institute of Biomedical Imaging and Bioengineering F32 Fellowship (EB022390). Data were provided in part by the HCP, MGH-USC Consortium (Principal Investigators: Bruce R. Rosen, Arthur W. Toga and Van Wedeen; U01MH093765), which was funded by the NIH Blueprint Initiative for Neuroscience Research grant; the National Institutes of Health grant P41EB015896; and the Instrumentation Grants S10RR023043, 1S10RR023401, 1S10RR019307.

Author Contributions B.Z., J.Z.L., S.F.C., B.R.R. and M.S.R. conceptualized the problem and contributed to experimental design. B.Z. developed, implemented and tested the technical framework. J.Z.L. and B.Z. constructed the theoretical description. B.Z., J.Z.L., S.F.C., B.R.R. and M.S.R. wrote the manuscript.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Publisher's note: Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations. Correspondence and requests for materials should be addressed to M.S.R. (mrosen@cfa.harvard.edu).

METHODS

Image dataset acquisition and pre-processing. The training dataset of generic images was assembled from ImageNet²⁰. 10,000 images from the 'animal', 'plant' and 'scene' categories were each cropped to the central 256×256 pixels and subsequently subsampled to 128×128 . Y-channel luminance was extracted from the RGB colour images to form greyscale intensity images. Each image was then rotated in 90° increments to augment the dataset. The mean intensity of each image was subtracted and the entire dataset was normalized to a constant value defined by the maximum intensity of the dataset.

The training dataset of de-identified brain images was assembled from the MGH-USC HCP¹⁹ public database, which were acquired with the T1-weighted three-dimensional MRI acquisition protocol MPRAGE with repetition time TR = 2,530 ms, echo time TE = 1.15 ms, inversion time TI = 1,100 ms, flip angle FA = 7.0° and bandwidth BW = 651 Hz Px^{-1} on a Siemens Skyra 3T MRI platform (Siemens Medical Solutions, Erlangen, Germany). The de-identification process used by the HCP protocol masked the face and ear regions to protect the subjects' privacy.

Axial, sagittal and coronal T1-weighted slices from 131 subjects were used to generate a 50,000-image dataset. For each image, the central 256×256 pixels were cropped and subsampled to 128×128 . To promote translation invariance in the training, each image was symmetrically tiled to create a larger 256×256 image containing four reflections of the original, and cropped to a random 128×128 section. The same data normalization process described above for the ImageNet dataset was used.

The test data used in the first evaluation experiment were taken from another subject outside those used for training. The test data also included the T2-weighted MRI acquisition protocol SPACE with TR = 3,200 ms, TE = 561 ms, FOV = $224 \text{ mm} \times 244 \text{ mm}$ and BW = 744 Hz Px^{-1} on the same Siemens Skyra 3T MRI platform.

Sensor-domain encoding of image data. Sensor-domain representations for each image were encoded according to the reconstruction task. For the Radon transform experiment (Fig. 2a–c), we used the discrete Radon transform with 180 projection angles and 185 parallel rays. The spiral k -space experiment (Fig. 2d–f) used nonuniform fast Fourier transform (NUFFT)³⁴ to encode a ten-interleave spiral trajectory³⁵ with variable density factor $\alpha = 1$ and undersampling factor $R = 1/1.2$ based on the pre-sampled 256×256 images (the MATLAB code used for this trajectory encoding is available at <http://bigwww.epfl.ch/algorithms/mri-reconstruction/>). The undersampled Cartesian k -space experiment (Fig. 2g–i) used a Poisson-disk sampling pattern with 40% undersampling of the Fourier-transformed k -space generated with the Berkeley Advanced Reconstruction Toolbox (BART)³⁶. The misaligned Cartesian k -space experiment (Fig. 2j–l) used Fourier-transformed k -space where an empirically observed phase shift from a typical echo planar imaging acquisition was applied to every odd readout line.

Modulation with synthesized phase. Phase-modulated training data was used in the raw sensor data experiments of Figs 4 and 5 to train networks to accurately process phase-modulated sensor data. Synthesized phase maps were created by generating two-dimensional sinusoids with varying spatial frequencies independent along each image axis, and rotated by a random angle with respect to the image axes. The intensities of the sinusoids represented phase values, and were normalized to be between 0 and 2π . Each magnitude image in the training dataset was then modulated with a randomly generated phase map to form the complex-valued target image, which was then encoded by the appropriate forward encoding model to produce the corresponding sensor-domain input.

Model architecture. The input to the neural network consists of a vector of sensor-domain-sampled data produced by the preprocessing steps detailed above. Because the input layer is fully connected to the first hidden layer, for each reconstruction task the sensor-domain data (typically represented in two dimensions for images) can be vectorized in any order without any effect on the training. Since the neural network computational framework used here (Tensorflow³⁷) operates on real-valued inputs and parameters, complex data must be separated into real and imaginary components concatenated in the input vector. Thus, an $n \times n$ complex-valued k -space matrix, for example, is reshaped to a $2n^2 \times 1$ real-valued vector (for our experiments, $n = 128$). As schematically illustrated in Fig. 1c, the input layer FC1 is fully connected to an $n^2 \times 1$ -dimensional hidden layer FC2 and activated by the hyperbolic tangent function. This first hidden layer is fully connected to another $n^2 \times 1$ -dimensional hidden layer FC3 with hyperbolic tangent activation, and is reshaped to an $n \times n$ matrix in preparation for convolutional processing. The first convolutional layer C1 convolves 64 filters of 5×5 with stride 1 followed by a rectifier nonlinearity³⁸. The second convolutional layer C2 again convolves 64 filters of 5×5 with stride 1 followed by a rectifier nonlinearity. The final output layer deconvolves the C2 layer with 64 filters of 7×7 with stride 1. The output layer represents the reconstructed magnitude image, except for the

phase-modulation experiment, where the network was trained separately to reconstruct the real and imaginary components of the image.

Training details. The same network architecture and hyperparameters were used for our experiments. For each sensor encoding reconstruction task, a different network was trained from the corresponding sensor-domain encodings and target images applied to the inputs and outputs, respectively, of the neural network (details of training data and network architecture described above). One per cent multiplicative noise was applied to the input to promote manifold learning during training by forcing the network to learn robust representations from corrupted inputs¹⁰. We note that the specific noise distribution of this corruption process did not serve to model the additive Gaussian noise that was applied during evaluation. The RMSProp algorithm (see http://www.cs.toronto.edu/~tijmen/csc321/slides/lecture_slides_lec6.pdf) was used with minibatches of size 100, learning rate 0.00002, momentum 0.0 and decay 0.9. The loss function minimized during training was a simple squared loss between the network output and target image intensity values, with an additional L1-norm penalty ($\lambda = 0.0001$) applied to the feature map activations in the final hidden layer C2 to promote sparse convolutional representations. The convolutional layers are inspired by Winner-Take-All autoencoders³⁹ that jointly optimize the sparse convolutional codes as well as the deconvolutional kernel 'dictionaries' upon which the final image is built (Fig. 3h). Note that this imposed sparsity on the convolutional layers is separate from the fully connected hidden-layer activation sparsity that emerged without an applied sparsifying penalty (Fig. 3a–c), and occurs even without imposed convolutional sparsity. Each network was trained for 100 epochs (duration typically 7–8 h) on the Tensorflow³⁷ deep learning framework using two NVIDIA Tesla P100 graphics processing units (GPUs) with 16-GB memory capacity each, specifically employing either a conventional server platform with two P100 GPUs or the NVIDIA DGX-1 using two GPUs per experiment. Example plots of training convergence are shown in Extended Data Fig. 3. The validation error tracking the training error without upward divergence demonstrates a stable training regime with good bias-variance tradeoff, indicating that model complexity is well matched to the reconstruction problem.

Evaluation on simulated sensor data. The performance of AUTOMAP-trained networks for the four acquisition strategies was evaluated by reconstructing the four sensor-domain encodings of T1- and T2-weighted MRI brain images of a human subject from the HCP database as described above. For the Radon transform, spiral k -space, and misaligned k -space experiments, the network was trained using ImageNet data; for the undersampled k -space experiment, the network was trained with data from the HCP brain image dataset using only T1-weighted images from other subjects in the HCP database.

We reconstructed the same set of sensor-domain inputs with conventional reconstruction techniques for each acquisition strategy: For Radon projection imaging, model-based iterative reconstruction⁴ was used with generalized Huber function parameters $\delta = 0.05$ and $T = 4.0$ and run until the average magnitude of voxel updates was less than 1%, implemented with OpenMBIR software available at <http://engineering.purdue.edu/~bouman/OpenMBIR/>. Spiral-trajectory k -space was reconstructed with a single-coil implementation of conjugate gradient-SENSE using NUFFT regridding^{5,34} with kernel size $J_d = 8$ samples, run over 30 conjugate gradient iterations, with MATLAB code available at <http://bigwww.epfl.ch/algorithms/mri-reconstruction/>. The Poisson-disk undersampled k -space was reconstructed with compressed sensing¹⁸ using the wavelet sparsifying transform with the L1 penalty parameter $\lambda = 0.01$, using BART³⁶ with code available at <https://mricon.github.io/bart/>. Misaligned k -space (commonplace sampling inaccuracy due to hardware limitations or physiologic effects) was reconstructed with the native MATLAB implementation of the two-dimensional inverse fast Fourier transform.

To probe the noise sensitivity of the reconstructions, varying levels of additive white Gaussian noise (AWGN) were introduced to the sensor-domain signals: 25 dB AWGN signal-to-noise ratio (SNR) for the spiral experiment, 30 dB AWGN SNR for the undersampled Cartesian experiment and 40 dB AWGN SNR for the Radon projection experiment; these SNR measures indicate the power level of the additive noise relative to the power of the signals. The influence of this additive noise on the resultant image SNR of the reconstructions was measured with a standard Monte Carlo SNR map calculation⁴⁰ $\text{SNR}(\mathbf{r}) = \bar{\mathbf{x}}_n(\mathbf{r}) / \sigma_n(\mathbf{r})$, where $\bar{\mathbf{x}}_n(\mathbf{r})$ and $\sigma_n(\mathbf{r})$ are the mean and standard deviation, respectively, over Monte Carlo instances $n = 1, 2, \dots, 100$, of the image magnitude at voxel \mathbf{r} . The representative SNR for an image was computed by taking the mean of the SNR map over the region of interest of voxels in the brain. We did not noise-corrupt the misaligned k -space because the sampling trajectory already represented a perturbed input. More extreme cases of noise corruption and its effects on reconstruction are shown in Extended Data Fig. 1.

Evaluation on raw MRI scanner data. Cartesian k -space test data (of Fig. 4) were acquired from a healthy volunteer on a 3T Siemens Trio MRI scanner with a spin-

echo imaging sequence with TR = 3,110 ms, TE = 23.0 ms, matrix size = 208 × 256 and slice thickness 3 mm. Data from the 12-channel receiver head coil was coil-compressed to one channel with singular value decomposition (SVD) and the central 128 × 128 k -space samples formed the input for the 128 × 128 matrix reconstruction task. As described above, the AUTOMAP network was trained on phase-modulated HCP brain data with Cartesian Fourier encoding; the acquired raw scanner data was then input to the trained network for reconstruction.

Spiral k -space data (Fig. 5) were acquired from a healthy volunteer on a Siemens 3T Prisma MRI scanner using a 32-channel head coil. A constant spiral trajectory was designed to cover a field of view of 256 × 256 mm² with 2 × 2 mm² in-plane resolution and 5 mm slice thickness. This was achieved using ten spiral interleaves, each having an 8-ms readout duration which included a 1-ms rewind, with slow rate 133 mT m⁻¹ ms⁻¹ and maximum gradient strength 24 mT m⁻¹; TE = 35 ms, TR = 200 ms, and flip angle of 20°. Data from the multichannel receiver head coil was SVD coil-compressed to one channel. A calibration acquisition measurement was made to measure the actual sampling trajectory⁴¹. The AUTOMAP network was trained on phase-modulated HCP brain data with non-Cartesian encoding with the designed spiral trajectory; acquired raw scanner data was then input to the trained network for reconstruction, and compared with NUFFT regridding reconstruction³⁴ with kernel size $f_d = 6$ samples, using code available from <http://web.eecs.umich.edu/~fessler/irt/irt/>.

Multichannel T2-weighted data (of Fig. 5) were acquired on a 3T Siemens Trio with the standard Siemens 32-channel head array coil. A turbo spin echo sequence with 224 × 224 mm² field of view was acquired across 35 slices with a 30% distance factor. The imaging parameters are as follows: TR = 6.1 s, TE = 98 ms, flip angle 150°, and a resolution of 0.5 × 0.5 × 3.0 mm³ with a matrix size of 448 × 448. The fully sampled uncombined complex k -space data were retrospectively undersampled to a 112 × 112 matrix, corresponding to 2 mm in-plane resolution. The channel data were then mixed down to 16 modes using the standard global SVD-based compression. Iterative SENSE reconstruction⁴² was performed using the GMRES solver⁴³ with a stopping criterion of 1×10^{-4} relative error to generate the ground truth reconstruction. Sensor-domain data were then undersampled by 15.5 times with an $R = 4 \times 4$ coherent undersampling pattern and 5 × 5 low-frequency region, and reconstructed with AUTOMAP and SENSE using the SVD coil sensitivity profiles. The AUTOMAP network was trained on HCP brain images, which were modulated by the SVD coil sensitivity profiles to produce the multichannel training data. Each channel was Fourier-transformed and correspondingly undersampled with the same $R = 15.5$ pattern, and channels were concatenated at the network input.

Evaluation on raw PET scanner data. PET data were acquired from a healthy volunteer using the Biograph mMR scanner (Siemens Healthineers, Erlangen, Germany). The emission data corresponding to the 45–60-min interval post-administration of about 5 mCi of ¹⁸F-FDG were used in this work. A PET volume was reconstructed using filtered back projection (Extended Data Fig. 5b) and the standard ordinary Poisson ordered subsets expectation maximization (OP-OSEM) algorithm⁴⁴ (Extended Data Fig. 5c), accounting for variable detector efficiency and photon attenuation and scatter using software provided by the manufacturer. The head attenuation map was generated from the magnetic resonance data using software developed in-house⁴⁵. Spatial smoothing was performed after image reconstruction using a 5-mm full-width at half-maximum (FWHM) Gaussian kernel. A set of attenuation-corrected two-dimensional sinograms corresponding to the direct planes was also generated from the three-dimensional sinograms using the single slice rebinning algorithm, and was used as input to the AUTOMAP reconstruction network, which was trained on T1-weighted brain images from the HCP database, encoded with discrete Radon transform and Poisson sampling using native MATLAB functions. Although the absence of a ground truth image makes it difficult to evaluate the differences between the reconstruction techniques (Extended Data Fig. 5b–d), this experiment demonstrates the ability of AUTOMAP to reconstruct PET data acquired on a human scanner with results comparable to clinically used reconstruction methods.

t-SNE analysis. Relationship of trained network weights were visualized with t-SNE²⁸. We employed a standard Cartesian Fourier k -space encoding for the networks. To reduce computational load, lower-resolution reconstruction networks were trained using 64 × 64 images from either ImageNet, brain images, or random-valued noise without any real-world image structure. In the visualization, each point corresponds to a single pixel in FC3, represented by an n^2 -dimensional vector of weights directed to it from the FC2 layer. The label for each point is a scalar pixel location in the image space (from 0 to n^2) that also defines its colour in the visualization; similar colours correspond to similar pixel location. The t-SNE algorithm was implemented with perplexity 64 over 200 iterations with MATLAB code available at <https://lvdmaaten.github.io/tsne/>.

Description of AUTOMAP manifold learning. Our learning task is twofold. Given \tilde{x} , the noisy observation of sensor-domain data x , we first want to learn the

stochastic projection operator onto \mathcal{X} : $p(\tilde{x}) = P(x|\tilde{x})$. After obtaining x , our second and more important task is to reconstruct $f(x)$ by producing a reconstruction mapping $\hat{f}: \mathbb{R}^{n^2} \rightarrow \mathbb{R}^{n^2}$ that minimizes the reconstruction error $L(\hat{f}(x), f(x))$.

We first describe the reconstruction process by considering the idealized scenario in which the input sensor data are noiseless. We denote the data $\{y_i, x_i\}_{i=1}^n$, where for the i th observation x_i indicates a $n \times n$ set of input parameters, and y_i indicates the $n \times n$ real, underlying images. We assume that (1) there exists a unknown smooth and homeomorphic function $f: \mathbb{R}^{n^2} \rightarrow \mathbb{R}^{n^2}$, such that $y = f(x)$, and (2) $\{x_i\}_{i=1}^n, \{y_i\}_{i=1}^n$ lie on unknown smooth manifolds \mathcal{X} and \mathcal{Y} , respectively. Both manifolds are embedded in the ambient space \mathbb{R}^{n^2} , such that $\dim(\mathcal{X}) < n^2$ and $\dim(\mathcal{Y}) < n^2$.

The above two assumptions combine to define a joint manifold $\mathcal{M}_{\mathcal{X},\mathcal{Y}} = \mathcal{X} \times \mathcal{Y}$ within which the entire dataset $(x_i, y_i)_{i=1}^n$ lies, which can be written as:

$$\mathcal{M}_{\mathcal{X},\mathcal{Y}} = \{(x, f(x)) \in \mathbb{R}^{n^2} \times \mathbb{R}^{n^2} | x \in \mathcal{X}, f(x) \in \mathcal{Y}\}$$

We note that as $(x, f(x))$ is described using the regular Euclidean coordinate system, we may equivalently describe this point using the intrinsic coordinate system of $\mathcal{M}_{\mathcal{X},\mathcal{Y}}$ as $(z, g(z))$, such that there exists a homeomorphic mapping $\phi = (\phi_x, \phi_y)$ between $(x, f(x))$ and $(z, g(z))$. That is, $x = \phi_x(z)$ and $f(x) = \phi_y \circ g(z)$. As an aside, in topology, $\phi = (\phi_x, \phi_y): \mathcal{M}_{\mathcal{X},\mathcal{Y}} \rightarrow \mathbb{R}^{n^2} \times \mathbb{R}^{n^2}$ corresponds to the local coordinate chart of $\mathcal{M}_{\mathcal{X},\mathcal{Y}}$ at the neighbourhood of $(x, f(x))$. Instead of learning f directly in the ambient space, we wish to learn the diffeomorphism g between \mathcal{X} and \mathcal{Y} in order to take advantage of the low-dimensional nature of the embedded space. Consequently, the process of generating $y = f(x)$ from x can be written as a sequence of function evaluations:

$$f(x) = \phi_y \circ g \circ \phi_x^{-1}(x)$$

For the convenience of later presentation, we notice that given input image x , the output image follows a probability distribution $Q(Y|X = x, f)$, which is a degenerate distribution with point mass at $y = f(x)$.

We now turn to the more realistic scenario where corruption exists in the sensor-domain input and describe the denoising process. Instead of observing the perfect input data x_i , we observe \tilde{x}_i , which is a corrupted version of x_i by some known corruption process described by the probability distribution $P(\tilde{X}|X = x)$. To handle this complication, we seek to learn a denoising step $Q(X|\tilde{X} = \tilde{x}, p)$ to our model pipeline, such that our prediction for y is no longer a deterministic value, but a random variable with conditional distribution $P(Y|\tilde{X})$ so that we can properly characterize the prediction uncertainty caused by the corruption process.

Instead of learning this denoising step explicitly, we draw an analogy from the denoising autoencoder and model the joint distribution $P(Y, X, \tilde{X})$ instead. Specifically, in addition to the assumptions (1) and (2) listed above, we also assume (3) that the true distribution $P(X|\tilde{X})$ lies in the semiparametric family \mathbb{Q} defined by its first moment $\mathbb{Q} = \{Q(X|\tilde{X} = \tilde{x}, p) | E(X) = p(\tilde{X})\}$.

We model $P(Y, X, \tilde{X})$ using the decomposition below:

$$Q_{(f,p)}(Y, X, \tilde{X}) = Q(Y|X, f)Q(X|\tilde{X}, p)P(\tilde{X})$$

where $Q(Y|X, f)$ denotes the model for the reconstruction process that we described earlier, $Q(X|\tilde{X}, p)$ is the denoising operator that we seek to learn, and $P(\tilde{X})$ is the empirical distribution of corrupted images. We note that we can combine the models for denoising and reconstruction processes together by collapsing the first two terms on the right-hand side into one term, which gives:

$$Q_{(f,p)}(Y, X, \tilde{X}) = Q(Y, X|\tilde{X}, (f, p))P(\tilde{X})$$

We note that $Y = f(X)$ is a deterministic and homeomorphic mapping of X , so $Q(Y, X|\tilde{X}, (f, p)) = Q(Y|\tilde{X}, (f, p))$ is the predictive distribution of output image y given the noisy input \tilde{x} , which is exactly our estimator of interest. Consequently, the model can be written as:

$$Q_{(f,p)}(Y, X, \tilde{X}) = Q(Y|\tilde{X}, (f, p))P(\tilde{X})$$

This completes the definition of our model for the joint distribution.

In the actual training stage, we usually took advantage of the fact that perfect input images x are available, and train the model with \tilde{x} that we generated from $P(\tilde{X}|X = x)$. That is to say, the joint distribution of (Y, X, \tilde{X}) observed in training data admits the form:

$$P(Y, X, \tilde{X}) = P(Y|X)P(\tilde{X}|X)P(X)$$

The training proceeds by minimizing the Kullback–Liebler divergence between observed probability $P(Y, X, \tilde{X})$ and our model $Q(Y, X, \tilde{X})$:

$$\mathbb{D}_{\text{KL}}[P(Y, X, \tilde{X}) \| Q_{(f,p)}(Y, X, \tilde{X})]$$

with respect to the function-valued parameters (f, p) . As the Kullback–Liebler divergence converges towards 0, $Q(X|\tilde{X}, p)$ converges to $P(X|\tilde{X})$, the denoising projection, and at the same time $Q(Y|\tilde{X}, (f, p))$ converges to $P(Y|\tilde{X})$.

There exists a rich literature^{46–50} on explicitly learning the stochastic projection p , diffeomorphism g , and the local coordinate chart ϕ . However, we notice that since $(\phi_f, \phi_x, p, g) \in \mathcal{C}^\infty$ (where \mathcal{C}^∞ denotes the set of infinitely differentiable functions), $\hat{f} = \phi_f \circ g \circ \phi_x^{-1} \circ p$ as a whole is a continuously differentiable function on a compact subset of \mathbb{R}^n , and can therefore theoretically be approximated by the universal approximation theorem⁵¹.

Practical considerations. While the idealized conception of the universal approximation theorem requires an infinite number of hidden nodes to achieve a perfect representation of an arbitrary continuous function⁵², it has been demonstrated that practical neural network implementations with finite hidden layers achieve a bounded approximation error for functions on a compact set, which applies to all domain transform functions that govern contemporary imaging systems^{21,53,54}. Furthermore, the system representation for state-of-the-art image reconstructions are discrete models that can be described exactly by a fully connected network (for example, in the simplest case, an inverse discrete Fourier transform matrix is applied to Nyquist sampled data). In addition, many inverse encoding models for image reconstruction can be represented using limited support (for example, SENSE reconstruction of uniformly or CAIPIRINHA (controlled aliasing in parallel imaging results in higher acceleration) staggered undersampled data results in a block diagonal inverse encoding, TV smoothness constraint for compressed sensing (Poisson matrix) produces a semiseparable inverse encoding with linear complexity, and so on). Using a discrete/finite neural network, AUTOMAP aims to expand upon these analytic models for image reconstruction.

The features identified in the AUTOMAP training process are not generated to create maximal distinction between certain categories of objects or subjects, as is common for many image classification tasks. Instead, our loss function forces minimization of pixelwise error and thus prioritizes reconstruction accuracy. As a result, our features are constrained to serve that purpose (note the resemblance of the deconvolutional kernels of Fig. 3h to Gabor-like edge detectors). This training approach makes AUTOMAP much less likely to produce ‘hallucinatory behaviours’ that arise from the use of finely tuned category-specific feature representations (as exploited by Google’s DeepDream system⁵⁵).

Given knowable systematic defects in the acquisition chain that can be modelled in the signal encoding (for example, measurable magnetic resonance gradient timing delays^{56,57}), an appropriate training set can be generated to allow AUTOMAP to compensate these acquisition nonidealities without manually designed postprocessing requiring expert knowledge. However, as with other reconstruction methods, untrained and unaccounted-for acquisition errors (for example, subject motion or voltage or current spikes in hardware) will produce errors in the reconstruction with the current implementation of AUTOMAP. Ideally, to detect and compensate for these and other unpredictable or object-dependent artefacts such as X-ray scattering in computed tomography or chemical shift in MRI would probably require incorporating a more sophisticated discriminator network⁵⁸ into AUTOMAP, similar to those used in generative adversarial networks to quantitatively evaluate the quality of the reconstructed image and iteratively adapt the reconstruction process via modification of the AUTOMAP network weights to reduce artefacts.

However, in this current implementation, AUTOMAP reconstruction is stable in the presence of typical sampling trajectory variation during an MRI acquisition (Fig. 5a–f). Extending our evaluation to another class of encoding errors, we performed a simulated motion corruption experiment using the TAMER motion encoding model⁵⁹. For standard multi-shot MRI acquisitions, patient motion between each imaging shot creates discrepancies in the encoding that often result in large ringing and blurring artefacts. As can be seen in Fig. 3, AUTOMAP extracts important data interdependency relationships and it is important to examine the robustness of these under realistic patient imaging situations. Using T2-weighted data (from the acquisition associated with Fig. 5g) we have simulated motion corruption using a realistic motion trajectory that was measured during an fMRI scan of a patient with Alzheimer’s disease. Specifically, in-plane motion variation was applied between each imaging shot to create representative artefacts which would be seen during a 2D Turbo Spin Echo acquisition with turbo factor 4. The fully sampled data were processed using AUTOMAP and the standard Fourier transform reconstruction. As can be seen in Extended Data Fig. 4, AUTOMAP does not exhibit instability in the presence of data corruption owing to patient motion and shows comparable artefact level and structure to standard reconstructions.

The role of expert knowledge in the AUTOMAP reconstruction framework requires careful consideration, especially for future developments into extended

applications. In this paper, our emphasis on withholding domain-specific knowledge is to explore the extreme case where expert knowledge is removed from the system where possible; thus our results are representative of what is probably a conservative limit on performance gains of introducing machine learning into the domain transform problem; we expect the appropriate integration of human domain-expert knowledge to AUTOMAP to yield even greater performance.

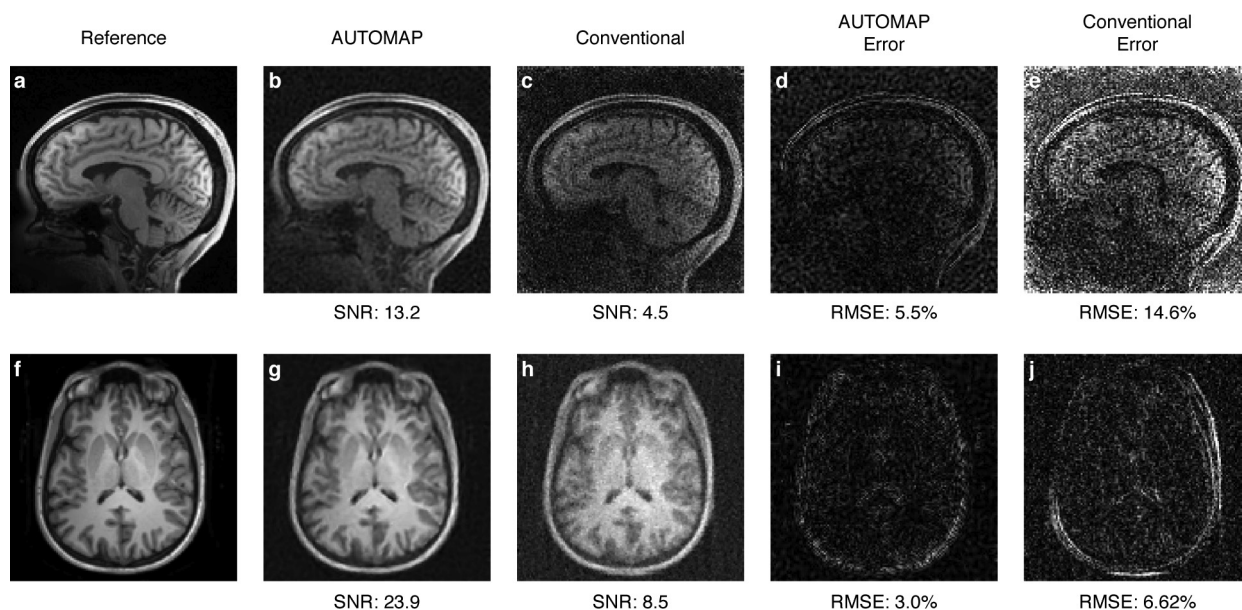
One particular area in which we anticipate benefits from expert knowledge is in the scaling of AUTOMAP to reconstruct higher-resolution images more efficiently. Given the current densely connected layer architecture, the number of weights in these layers scale by n^2 (compared to the fast Fourier transform complexity of the order of $n \log n$, for example), which can be problematic as memory and computation are limited resources. However, it is important to note the role that locality and separability play in image reconstruction. When considering standard MRI reconstruction methods, locality can be easily observed by examining the point spread function that defines the interaction between imaging voxels^{60–63}. This strong locality is an important property because it relates to the stability of solving the inverse problem, and has the added advantage of often resulting in nearly or fully decoupled reconstruction problems. As has been demonstrated in the recent MRI literature, decomposition methods such as the alternating direction method of multipliers (ADMM) have been used effectively to combine smaller semiseparable reconstruction problems to solve larger fully coupled optimizations (for example, maximum likelihood)^{64,65}. In the presence of hardware limitations, we expect the AUTOMAP framework to be directly employed on domain decomposed subsets to perform larger reconstructions. This would simply require the training of networks that describe the influence of subsets of acquired data to overlapping subsets of voxels. Given the success of methods that take advantage of locality properties (Schwarz alternating method/domain decomposition in linear algebra, Dantzig–Wolfe decomposition for linear programming, ADMM for convex optimization, and so on), we do not envision this being a limiting factor for AUTOMAP. In this light, the ability of AUTOMAP to accurately and robustly represent an inverse encoding model for the reconstruction of real-world multi-channel data (Fig. 5g–k) will allow it to serve as a core building block for large-scale reconstructions through the use of domain decomposition, variable splitting and alternating direction methods. Finally, clinically oriented implementations of AUTOMAP would probably be focused on a particular imaging modality or class of acquisitions within a modality, and would thus allow for limiting the universality of the network to obtain compact representations that can be generated through network compression methods such as pruning^{66,67}.

Data availability. The generic natural images used for training are available from the ImageNet database (<http://www.image-net.org/>). The brain images used for training and evaluation were obtained from the MGH-USC HCP database (<https://db.humanconnectome.org/>).

Code availability. Source code is available from the corresponding author upon reasonable request.

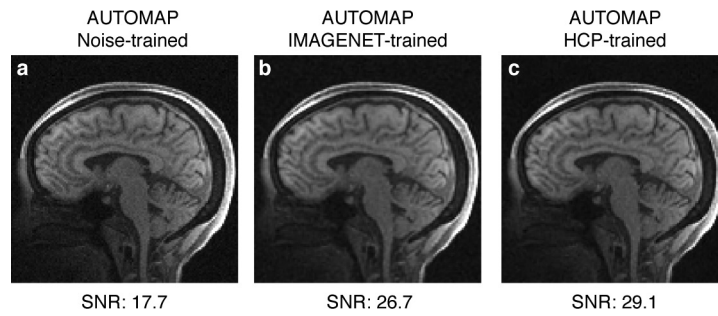
34. Fessler, J. A. & Sutton, B. P. Nonuniform fast Fourier transforms using min-max interpolation. *IEEE Trans. Signal Process.* **51**, 560–574 (2003).
35. Kim, D. H., Adalsteinsson, E. & Spielman, D. M. Simple analytic variable density spiral design. *Magn. Reson. Med.* **50**, 214–219 (2003).
36. Uecker, M., Ong, F., Tamir, J. I. & Bahri, D. Berkeley advanced reconstruction toolbox. *Proc. Int. Soc. Magnetic Resonance in Medicine* 2486 (2015).
37. Abadi, M. et al. TensorFlow: large-scale machine learning on heterogeneous distributed systems. Preprint at <https://arxiv.org/abs/1603.04467> (2016).
38. Nair, V. & Hinton, G. E. Rectified linear units improve restricted Boltzmann machines. In *Proc. 27th Int. Conf. on ‘Machine Learning’* 807–814 (ACM, 2010).
39. Makhzani, A. & Frey, B. J. Winner-take-all autoencoders. *Adv. Neural Inf. Process. Syst.* **28**, 2791–2799 (2015).
40. Reeder, S. B. et al. Practical approaches to the evaluation of signal-to-noise ratio performance with parallel imaging: application with cardiac imaging and a 32-channel cardiac coil. *Magn. Reson. Med.* **54**, 748–754 (2005).
41. Duyn, J. H., Yang, Y., Frank, J. A. & van der Veen, J. W. Simple correction method for k-space trajectory deviations in MRI. *J. Magn. Reson.* **132**, 150–153 (1998).
42. Pruessmann, K. P., Weiger, M., Scheidegger, M. B. & Boesiger, P. SENSE: sensitivity encoding for fast MRI. *Magn. Reson. Med.* **42**, 952–962 (1999).
43. Saad, Y. & Schultz, M. H. GMRES: a generalized minimal residual algorithm for solving nonsymmetric linear systems. *SIAM J. Sci. Statist. Comput.* **7**, 856–869 (1986).
44. Comtat, C. et al. OSEM-3D Reconstruction Strategies for the ECAT HRRT. *IEEE Symp. Conf. Record Nuclear Science* **6**, 3492–3496 (2004).
45. Izquierdo-Garcia, D. et al. An SPM8-based approach for attenuation correction combining segmentation and nonrigid template formation: application to simultaneous PET/MR brain imaging. *J. Nucl. Med.* **55**, 1825–1830 (2014).
46. Yu, K. & Zhang, T. Improved local coordinate coding using local tangents. In *Proc. 27th Int. Conf. on ‘Machine Learning’* 1215–1222 (ACM, 2010).

47. Anderes, E. & Coram, M. A general spline representation for nonparametric and semiparametric density estimates using diffeomorphisms. Preprint at <https://arxiv.org/abs/1205.5314> (2012).
48. Zhang, M., Singh, N. & Fletcher, P. T. Bayesian estimation of regularization and atlas building in diffeomorphic image registration. *Int. Conf. Inf. Process. Med. Imaging* 37–48 (Springer, 2013).
49. Fishbaugh, J., Prastawa, M., Gerig, G. & Durrleman, S. Geodesic image regression with a sparse parameterization of diffeomorphisms. In *1st Int. Conf. on 'Geometric Science of Information' GSI 2013* (eds Nielsen, F. & Barbaresco, F.) Vol. 8085, 95–102, https://link.springer.com/chapter/10.1007/978-3-642-40020-9_9 (2013).
50. Bernstein, A., Kuleshov, A. & Yanovich, Y. *Manifold Learning in Regression Tasks. Statistical Learning and Data Sciences* 414–423 (Springer, 2015).
51. Hornik, K. Approximation capabilities of multilayer feedforward networks. *Neural Netw.* **4**, 251–257 (1991).
52. Irie, B. & Miyake, S. Capabilities of three-layered perceptrons. In *IEEE Int. Conf. on 'Neural Networks'* Vol. 1, 641–648 (1988).
53. Cybenko, G. Approximation by superpositions of a sigmoidal function. *Math. Contr. Signals Syst.* **2**, 303–314 (1989).
54. Barron, A. R. Approximation and estimation bounds for artificial neural networks. *Mach. Learn.* **14**, 115–133 (1994).
55. Mordvintsev, A., Olah, C. & Tyka, M. DeepDream—a code example for visualizing neural networks. <https://research.googleblog.com/2015/07/deepdream-code-example-for-visualizing.html> (Google Res, 2015).
56. Addy, N. O., Wu, H. H. & Nishimura, D. G. Simple method for MR gradient system characterization and k-space trajectory estimation. *Magn. Reson. Med.* **68**, 120–129 (2012).
57. Han, H., Ouriadov, A. V., Fordham, E. & Balcom, B. J. Direct measurement of magnetic field gradient waveforms. *Concepts Magn. Reson.* **36A**, 349–360 (2010).
58. Goodfellow, I., Pouget-Abadie, J. & Mirza, M. Generative adversarial nets. *Adv. Neural Inf. Process. Syst.* 2672–2680 (2014).
59. Haskell, M., Cauley, S. F. & Wald, L. L. Targeted Motion Estimation and Reduction (TAMER): data consistency based motion mitigation for MRI using a reduced model joint optimization. *IEEE Trans. Med. Imaging* PP, 99, <http://doi.org/10.1109/TMI.2018.2791482> (2018).
60. Fessler, J. A., Lee, S., Olafsson, V. T., Shi, H. R. & Noll, D. C. Toeplitz-based iterative image reconstruction for MRI with correction for magnetic field inhomogeneity. *IEEE Trans. Signal Process.* **53**, 3393–3402 (2005).
61. Cauley, S. F. *et al.* Fast reconstruction for multichannel compressed sensing using a hierarchically semiseparable solver. *Magn. Reson. Med.* **73**, 1034–1040 (2015).
62. Xi, Y., Xia, J., Cauley, S. & Balakrishnan, V. Superfast and stable structured solvers for Toeplitz least squares via randomized sampling. *SIAM J. Matrix Anal. Appl.* **35**, 44–72 (2014).
63. Xia, J., Chandrasekaran, S., Gu, M. & Li, X. S. Fast algorithms for hierarchically semiseparable matrices. *Numer. Linear Algebra Appl.* **17**, 953–976 (2010).
64. Weller, D. S., Ramani, S. & Fessler, J. A. Augmented Lagrangian with variable splitting for faster non-Cartesian L1-SPIRiT MR image reconstruction. *IEEE Trans. Med. Imaging* **33**, 351–361 (2014).
65. Zhao, B., Setsompop, K., Ye, H., Cauley, S. F. & Wald, L. L. Maximum likelihood reconstruction for magnetic resonance fingerprinting. *IEEE Trans. Med. Imaging* **35**, 1812–1823 (2016).
66. Han, S., Mao, H. & Dally, W. J. Deep compression: compressing deep neural networks with pruning, trained quantization and Huffman coding. Preprint at <https://arxiv.org/abs/1510.00149> (2015).
67. Hu, H., Peng, R., Tai, Y.-W. & Tang, C.-K. Network trimming: a data-driven neuron pruning approach towards efficient deep architectures. Preprint at <https://arxiv.org/abs/1607.03250> (2016).



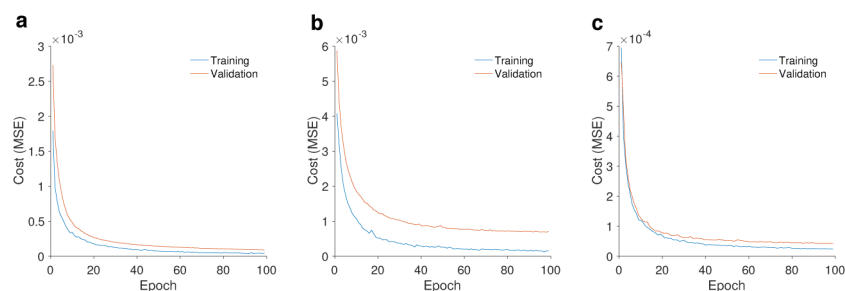
Extended Data Figure 1 | Reconstruction performance of AUTOMAP in low-signal-to-noise-ratio regimes. Reference brain images were encoded into sensor-domain sampling strategies with high levels of additive white Gaussian noise and reconstructed using both AUTOMAP and conventional approaches: **a–e**, spiral k -space encoding compared with conjugate-gradient SENSE reconstruction with NUFFT regridding;

f–j, Radon projection encoding compared with model-based iterative reconstruction. Image magnitude signal-to-noise ratios (SNRs) and error maps (with root mean squared error calculations) with respect to reference ground truth images are also shown. For each encoding experiment, both error maps are windowed to the same level.



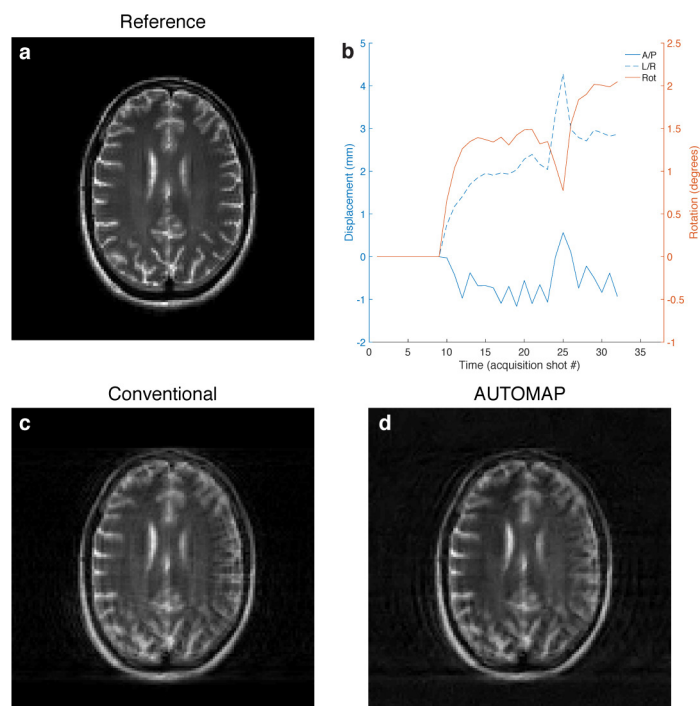
Extended Data Figure 2 | Effect of training corpus on image reconstruction. **a–c**, AUTOMAP was trained using sensor-image pairs of Cartesian Fourier encoded corpora derived from either ImageNet, HCP brain images, or random-valued Gaussian noise without any real-world image structure. Each trained network was then used to reconstruct a

noise-corrupted Cartesian k -space brain dataset. The signal-to-noise ratio (SNR) of the reconstructed images is shown. The apparent intensity discontinuity in the region above the eyes is due to the masking process used to de-identify the data in the HCP protocol (see Methods for more details).



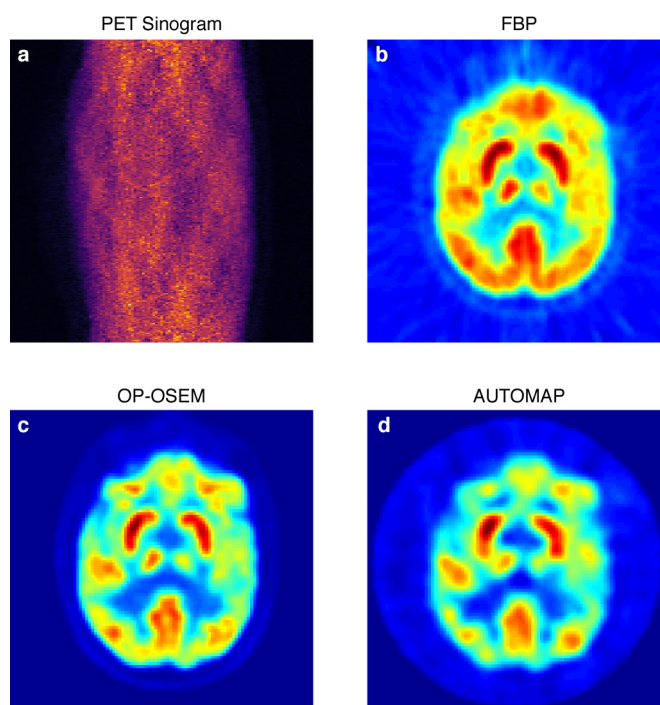
Extended Data Figure 3 | Training curves of optimizer loss convergence. Mean squared error (MSE) loss was minimized with stochastic gradient descent using the RMSProp algorithm and plotted here against training epoch count for: **a**, Cartesian Fourier encoding on IMAGENET corpus;

b, spiral Fourier encoding on IMAGENET corpus; and **c**, Cartesian undersampled Fourier encoding on HCP brain corpus. The validation error tracks the training error without upward divergence, demonstrating a stable training regime with good bias-variance tradeoff.



Extended Data Figure 4 | Reconstruction of motion-corrupted MRI. **a**, T2-weighted reference image acquired at 3 T with a turbo spin-echo sequence. **b**, Three-dimensional motion trajectories measured during an Alzheimer's patient study. **c**, **d**, These motion trajectories were used to corrupt the k -space of this reference image, and it was reconstructed without motion compensation using inverse Fourier transform (**c**)

and AUTOMAP (**d**). Both images show comparable artefact level and structure, demonstrating the stability of AUTOMAP reconstruction in the presence of unanticipated subject motion. A/P refers to anterior and posterior translational motion, L/R refers to left and right translational motion.



Extended Data Figure 5 | Reconstruction of PET scanner data. a–d, Human FDG PET sinogram data (a) was reconstructed using (b) filtered back projection (FBP), (c) OP-OSEM and (d) AUTOMAP.

Continuous-wave room-temperature diamond maser

Jonathan D. Breeze^{1,2}, Enrico Salvadori^{3,4,5}, Juna Sathian¹, Neil McN. Alford^{1,2} & Christopher W. M. Kay^{3,4,6}

The maser—the microwave progenitor of the optical laser—has been confined to relative obscurity owing to its reliance on cryogenic refrigeration and high-vacuum systems. Despite this, it has found application in deep-space communications and radio astronomy owing to its unparalleled performance as a low-noise amplifier and oscillator. The recent demonstration of a room-temperature solid-state maser that utilizes polarized electron populations within the triplet states of photo-excited pentacene molecules in a *p*-terphenyl host^{1–3} paves the way for a new class of maser. However, *p*-terphenyl has poor thermal and mechanical properties, and the decay rates of the triplet sublevel of pentacene mean that only pulsed maser operation has been observed in this system. Alternative materials are therefore required to achieve continuous emission: inorganic materials that contain spin defects, such as diamond^{4–6} and silicon carbide⁷, have been proposed. Here we report a continuous-wave room-temperature maser oscillator using optically pumped nitrogen–vacancy defect centres in diamond. This demonstration highlights the potential of room-temperature solid-state masers for use in a new generation of microwave devices that could find application in medicine, security, sensing and quantum technologies.

Solid-state masers, developed in the 1960s, were realized by pumping the spin states of paramagnetic impurities, such as Cr³⁺ ions doped into single-crystal sapphire (ruby). Pumping three-level systems with microwaves generates a population inversion—the requisite feature for amplification by stimulated emission in both masers and lasers—whereby a higher-energy state is more populated than a lower-energy state. For these early paramagnetic systems, cryogenic cooling was necessary to reduce the effects of spin–lattice relaxation, prolonging the lifetime of the population inversion and making continuous maser emission feasible.

The first demonstration of a room-temperature solid-state maser required more than 200 W of optical power to overcome the masing threshold, and produced a burst of microwave power at 1.45 GHz (ref. 1). This maser was subsequently improved and miniaturized using strontium titanate as the dielectric resonator material^{2,3}, which lowered the optical pump threshold and volume by two orders of magnitude. However, the threshold pump rate per molecule remained the same at about 10⁴ s^{−1}. Even if continuous-wave operation could be achieved in organic triplet-based masers, heating is a problem because most of the absorbed optical pump power is dissipated through non-radiative relaxation processes and the organic hosts have poor thermal properties. For example, the pentacene host *p*-terphenyl has a low thermal conductivity of 0.1 W m^{−1} K^{−1} and a melting point of 230 °C.

Inorganic materials have better thermal and mechanical properties than do organic materials. Specifically, those that host spin defects offer a potential route to realizing continuous-wave room-temperature solid-state masers if the population of defect spin states can be polarized (inverted). Silicon carbide (SiC) is a promising candidate because its various types of spin defect can be addressed individually by tuning the wavelength of optical excitation and applying a magnetic field, which,

through the Zeeman interaction, can select specific spin multiplicities. Silicon–vacancy defects have spin-quadruplet ground states which can be polarized through optical pumping. Continuous stimulated emission of microwaves from optically pumped silicon–vacancy defects in 6H-SiC has been reported recently⁷, but masing remained elusive.

In the 1960s, nitrogen impurities in diamond were proposed as potential maser gain media^{8–10}, by using a four-spin-flip cross-relaxation mechanism to produce population inversion in paramagnetic spin-1/2 nitrogen donors. Charged nitrogen–vacancy (NV[−]) defect centres in diamond (Fig. 1a) were observed using electron paramagnetic resonance (EPR) almost 40 years ago¹¹. The ability to prepare and read the quantum state of the triplet sublevels efficiently using optically detected magnetic resonance has enabled applications of NV[−] centres in magnetometry^{12–14} and quantum information processing¹⁵. They were recently proposed as quantum emitters for room-temperature masers⁶ owing to their attractive properties of long spin dephasing times (of greater than 1 μs), long spin-polarization lifetimes (of about 5 ms; refs 16, 17) and a triplet ground state that can be polarized through optical pumping^{6,18,19}. Furthermore, diamond has the highest known thermal conductivity (10³ W m^{−1} K^{−1}) and excellent mechanical properties, which obviates thermal runaway. Here we use ensembles of optically pumped NV[−] centres as the gain medium for a continuous-wave room-temperature maser oscillator.

The triplet ground state (³A₂) of the NV[−] defect has a zero-field splitting *D* that places the quasi-degenerate |±1⟩ sublevels approximately 2.87 GHz above the |0⟩ sublevel (Fig. 1b)¹⁹. At room temperature and in zero magnetic field, the sublevels are populated according to Boltzmann statistics, with a few more electrons populating the lower |0⟩ sublevel than the higher |±1⟩ sublevels. The triplet ground state can be spin-polarized by pumping with optical radiation with wavelengths of around 532 nm (ref. 18), so that the |0⟩ sublevel becomes preferentially populated with respect to |±1⟩. On photo-excitation, electrons in the |0⟩ and |±1⟩ sublevels of the triplet ground state undergo spin-conserving transitions into an excited triplet state (³E). Spin-selective intersystem crossing preferentially transfers electrons from the |±1⟩ excited triplet-state sublevels to a metastable singlet state, which then decay non-radiatively back to the ground state, into the |0⟩ and |±1⟩ sublevels at roughly equal rates. In this manner, continuous optical pumping of the triplet ground state of the NV[−] defect can result in spin polarization whereby up to 80% of electrons reside in the |0⟩ sublevel¹⁹.

However, stimulated emission requires a positive population inversion. This can be achieved by applying a magnetic field along one of the four axes of the NV defects (Fig. 1a), which splits the energy of the |±1⟩ sublevels via the Zeeman interaction but leaves the |0⟩ sublevel unchanged (Fig. 1c). For an electron Zeeman interaction energy of $\gamma_e \hbar B$, greater than twice the zero-field-splitting energy $\hbar D$ (for $B \geq 102.5$ mT), the energy of the |−1⟩ sublevel dips below that of the |0⟩ sublevel, permitting a population inversion to be established. If the transition frequency ω_s between the ground state |*g*⟩ ≡ |−1⟩ and the excited state |*e*⟩ ≡ |0⟩ is resonant with that of a cavity mode ω_c , then

¹Department of Materials, Imperial College London, Exhibition Road, London SW7 2AZ, UK. ²London Centre for Nanotechnology, Imperial College London, Exhibition Road, London SW7 2AZ, UK.

³Institute of Structural and Molecular Biology, University College London, Gower Street, London WC1E 8BT, UK. ⁴London Centre for Nanotechnology, 17–19 Gordon Street, London WC1H 0AH, UK.

⁵School of Biological and Chemical Sciences, Queen Mary University of London, Mile End Road, London E1 4NS, UK. ⁶Department of Chemistry, University of Saarland, 66123 Saarbrücken, Germany.

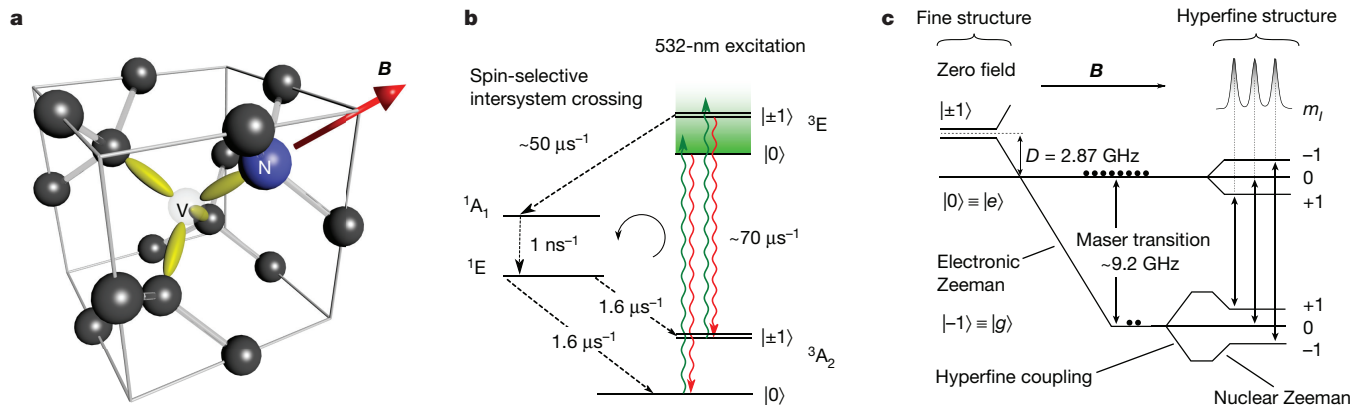


Figure 1 | Structure of the NV⁻ centres, optical pumping and magnetic-field interaction. **a**, The NV⁻ centre in diamond comprises a nitrogen atom (blue) adjacent to a vacancy (light grey), surrounded by carbon atoms (black). A magnetic field **B** (red arrow) is applied along the NV axis. **b**, The optical spin-polarizing pump process at zero-field. A continuous-wave 532-nm laser pumps electrons from the ³A₂ triplet ground state into the triplet excited state ³E via spin-conserving radiative transitions. Non-radiative spin-selective intersystem crossing transfers roughly 40% of the excited $|\pm 1\rangle$ electrons into the singlet state ¹A₁ at a rate of 50 μs⁻¹, which decay quickly (at a rate of 1 ns⁻¹) into the metastable state ¹E, and then back to the triplet ground state with roughly equal probability of entering the $|\pm 1\rangle$ or $|0\rangle$ sublevels (at a rate 1.6 μs⁻¹). The remaining electrons in the excited triplet state fluoresce back to the ground state

stimulated emission of microwave photons will occur. Furthermore, if the rate of stimulated emission of microwave photons exceeds the loss rate due to cavity dissipation, then build-up of the cavity-photon population occurs, resulting in maser oscillation.

For this study, we used a synthetic type-IIa diamond with a concentration of NV⁻ centres of 0.36 p.p.m. (Methods). The number of NV⁻ centres was estimated to be 7.3×10^{14} , of which roughly one in eighteen couple to the cavity mode, as a result of two of the three triplet sublevels being resonant on one of the four NV axes and one of the three ¹⁴N hyperfine lines, yielding $N = 4.0 \times 10^{13}$ active NV⁻ centres. The longitudinal (spin-lattice) and transverse (spin-dephasing) relaxation times were measured at 9.5 GHz and room temperature (300 K) using EPR spectroscopy, yielding $T_1 = 4.8$ ms and $T_2^* = 0.5$ μ s, respectively (Methods). These values agree with those reported for samples with similar concentrations of nitrogen impurities^{16,17}.

The cooperativity $C = 4g_s^2 N / (\kappa_c \kappa_s)$ is a figure of merit for masers and lasers, where g_s is the single spin-photon coupling rate, $\kappa_c = \omega_c / Q$ is the decay rate of the cavity mode, Q is the quality factor of the cavity and $\kappa_s \approx 2/T_2^*$ the spin-dephasing rate. For above-threshold maser oscillation the cooperativity needs to be much greater than unity ($C \gg 1$). The Purcell factor is key to enhancing maser performance^{2,20} and contributes a factor of g_s^2 / κ_c to the cooperativity. A maser cavity with a high Purcell factor was designed and constructed using a hollow cylindrical single-crystal sapphire ring housed within a copper cavity. The cavity supports an axisymmetric transverse electric ($TE_{01\delta}$) mode that resonates at 9.22 GHz with a loaded Q factor of 30,000. A magnetic mode volume of $V_m = 0.15 \text{ cm}^3$ was calculated, yielding a single spin-photon coupling of $g_s = 0.7 \text{ Hz}$ and an estimated (\sqrt{N} -enhanced) collective spin-photon coupling of $g_e = 4.4 \text{ MHz}$ (Methods). This collective coupling, combined with the decay rate of the cavity mode $\kappa_c \approx 1.9 \text{ MHz}$ and the measured spin-decoherence rate $\kappa_s \approx 2/T_2^* = 3.9 \text{ MHz}$, yields a cooperativity of $C \approx 10.6$, which indicates that the maser should oscillate above a threshold pump rate per active NV^- centre of $w_{\text{thr}} \approx 300 \text{ s}^{-1}$, equivalent to about 180 mW of optical power.

The cavity was positioned between the poles of an electromagnet (Bruker; Fig. 2) and oriented horizontally to allow optical access to continuous 532-nm laser excitation. The diamond was placed inside

by spin-conserving radiative processes at a rate of $70 \mu\text{s}^{-1}$. This process preferentially populates the $|0\rangle$ sublevel, producing a spin-polarized triplet ground state^{6,18,19}. c, The Zeeman and hyperfine interactions of NV centres. A magnetic field \mathbf{B} applied parallel to the NV⁻ axis splits the $|\pm 1\rangle$ states through the electronic and nuclear Zeeman interaction. For fields strengths of more than about 102.5 mT, the energy of the $|-1\rangle$ state drops below that of the $|0\rangle$ state, the difference at zero field being $D \approx 2.87$ GHz. If the $|0\rangle$ triplet sublevels are preferentially populated through optical pumping, then population inversion (with electrons depicted as black circles) is established. Hyperfine coupling between ($S = 1$) spin triplets with adjacent ^{14}N nuclear spins m_I (where $I = 1$) produces three observable emission lines. The maser transition is between the $|e\rangle \equiv |0\rangle$ and $|g\rangle \equiv |-1\rangle$ states.

the sapphire ring and, using EPR spectroscopy, oriented so that one set of NV $\langle 111 \rangle$ axes was aligned parallel to the magnetic field (Methods, Extended Data Fig. 1). EPR spectra were collected in the dark (Fig. 3a, b) and then laser excitation was applied with gradually increasing power. As the laser power increases, the high-field absorption lines decrease in amplitude until they disappear at a pump power of about 1.5 mW, owing to equalization of the $|e\rangle$ and $|g\rangle$ populations (zero spin

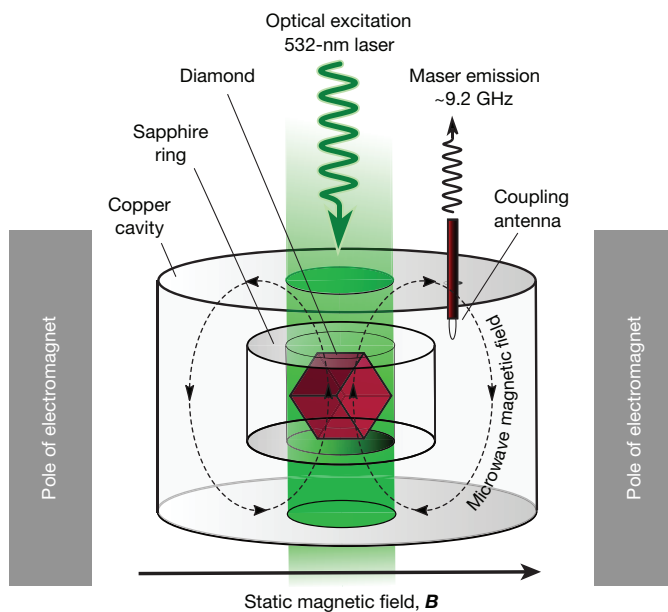


Figure 2 | Diamond maser construction. The diamond was placed inside the sapphire ring and, using EPR spectroscopy, oriented so that one set of NV(111) axes was aligned parallel to the magnetic field (Methods, Extended Data Fig. 1). The sapphire and diamond were housed within a cylindrical copper cavity, and a microwave loop antenna coupled the maser emission power into a transmission line. The diamond was optically pumped at 532 nm by a continuous-wave Nd:YAG laser.

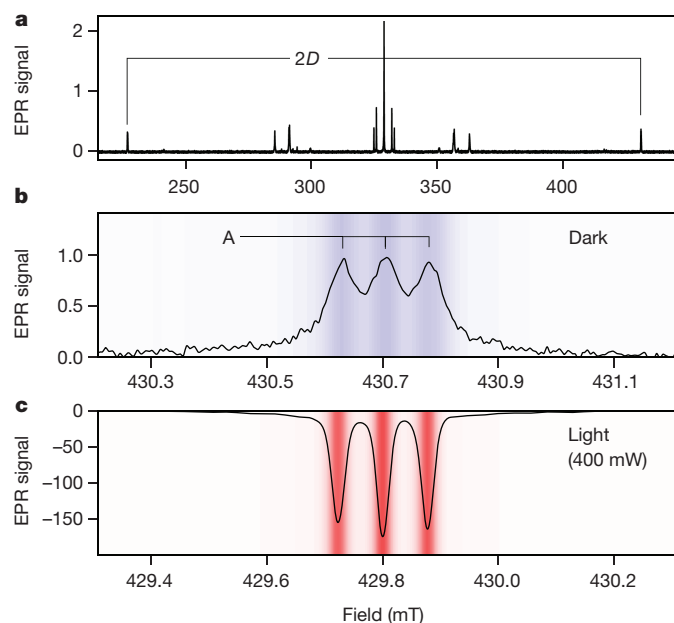


Figure 3 | Electron paramagnetic resonance spectroscopy. **a**, Wide EPR field sweep from 215 mT to 445 mT. The diamond sample was aligned to provide a maximum field splitting of $2D \approx 102.5$ mT between the low-field and high-field lines. **b**, High-field absorption (blue shading) in the dark state, normalized to the maximum value. The three lines are separated by approximately $A = 0.075$ mT, owing to the hyperfine coupling of ($S = 1$) triplets with adjacent ($I = 1$) ^{14}N nuclei. **c**, The lines become emissive (red shading) upon illumination by 400 mW of 532-nm laser light. The EPR signal amplitude is roughly 150 times greater than in the dark state and has shifted -25 MHz in frequency (about 0.9 mT).

polarization). Modelling of the populations under optical pumping yields an equalization pump rate of $w_{\text{eq}} = 1.5 \text{ s}^{-1}$, corresponding to an estimated pump power of about 0.9 mW, which agrees well with the measured incident power. The low concentration of NV centres results in only 33% of the incident optical power being absorbed. Although this is far from optimal, it ensures that the NV centres are pumped almost homogeneously, because the laser intensity varies marginally across the sample.

When the laser pump power was increased to 400 mW (Fig. 3c), we observed a cavity frequency shift of -25 MHz. We attribute this shift to an estimated 35°C increase in temperature of the sapphire ring and diamond, caused by heating. The observed EPR signal was a factor of about 150 greater in amplitude than for the dark state. Because the amplitude of the EPR signal is proportional to the difference in the populations of the $|e\rangle$ and $|g\rangle$ states (the spin polarization), it can be calibrated against the Boltzmann-populated dark state to provide an estimate of the population inversion, $S^z = N_e - N_g \approx 0.09N$, which agrees favourably with the predicted inversion⁶, $S^z = \kappa_c \kappa_s / (4g_s^2) \approx 0.10N$. Once the optical beam alignment and polarization were optimized for maximum emission on the high-field line and the cavity frequency had stabilized, the maser was disconnected from the EPR spectrometer and connected to a spectrum analyser (Advantest R3271A). The magnetic field was stepped across the high-field resonances using a programmable field controller (Bruker ER-032M) and the microwave output recorded at each magnetic field position. As depicted in Fig. 4a, three separate regions of continuous maser emission—corresponding to the three hyperfine transitions—were observed. The peak emission power of the maser was -90.3 dBm (comparable with a hydrogen maser), with an estimated average pump rate per NV^- centre of $w = 410 \text{ s}^{-1}$. The maser emission persisted without degradation in power for the duration of all experiments (the longest being 10 h), demonstrating the robustness of the system. We repeated the experiment for different laser

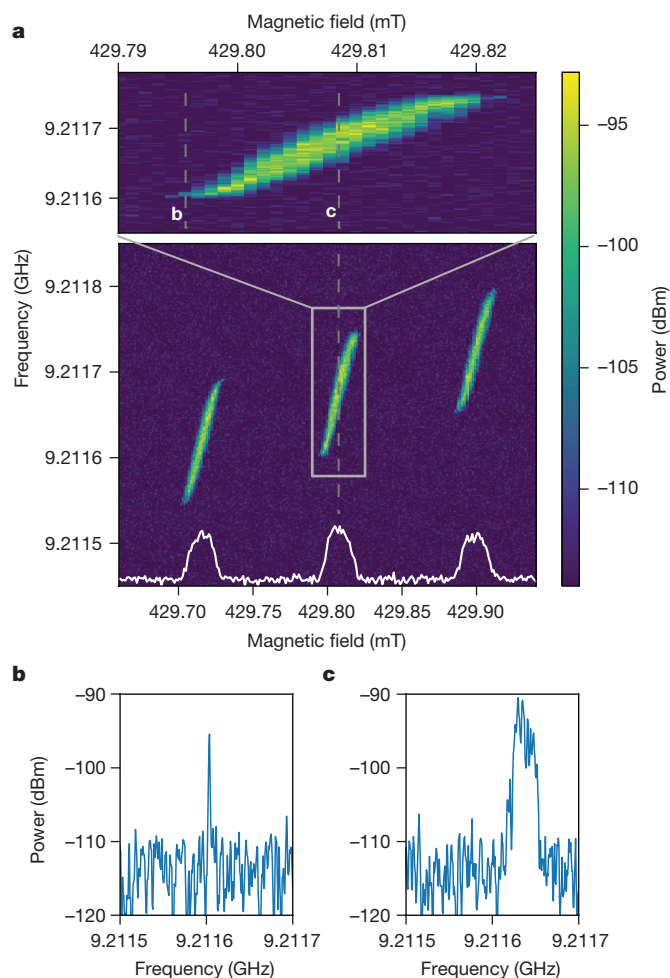


Figure 4 | Field-frequency maser emission plots. For an optical pump power of 400 mW, the static magnetic field B was varied from 429 mT to 431 mT in $1\text{-}\mu\text{T}$ increments. At each field value, we recorded the spectral emission power that emerged from the maser. **a**, The maser emission spectral power (colour scale) as a function of applied magnetic field B . There are three regions of maser oscillation, which correspond to each of the hyperfine transitions. The white line shows the integrated power spectral density. The central region is magnified (top) to show the narrow-linewidth emission spectrum at the edge of the threshold (**b**) and the limit-cycle-broadened spectrum at the centre of the emission region (**c**). **b**, At the edge of the emission spectrum, slightly above threshold the emission has a narrow linewidth (about 50 Hz). **c**, At the centre of the emission region, the linewidth is broadened owing to the presence of limit cycles.

pump powers, which revealed threshold behaviour with a threshold pump power of 138 mW (Extended Data Fig. 2a).

At the edge of the maser emission region (Fig. 4b), the linewidth of approximately 50 Hz approaches the Schawlow–Townes limit²¹ for the observed output power P_{out} , $\gamma_{\text{ST}} = \pi \hbar \omega_c \kappa_c^2 / (2P_{\text{out}}) \approx 10$ Hz. However, near the centre of the each emission region, the maser emission line appears broadened (Fig. 4c). Dynamical simulations of the photon population of the maser cavity reveal nonlinear oscillations that exhibit limit cycles²² in the inversion–photon–population phase space. Further investigation revealed that these limit cycles are eliminated if the spin decay rate κ_c is increased (by increasing the concentration of NV centres) or if the spin–photon coupling rate g_s is decreased through detuning ($\omega_c \neq \omega_s$), which explains their absence just above threshold (Fig. 4b).

We have achieved continuous maser operation by combining a cavity with a high Purcell factor with the narrow linewidth of the NV^- defect transitions in diamond. Although we used a frequency of 9.2 GHz in this study, which is close to the 9.193 GHz of a caesium

atomic clock, any frequency that is required may be produced by changing the magnetic field to be resonant at the appropriate frequency. We envisage that a diamond maser amplifier working at room temperature can find immediate application in deep-space communications and radio astronomy, where its ultralow noise would be beneficial in amplifying weak electromagnetic signals without the need for cryogenic cooling. Furthermore, ultrasensitive parametric detectors could be used in magnetic resonance, secure communications and high-precision metrology. Given that the fundamental process that occurs in a maser is the conversion of optical photons to coherent microwave photons²⁰, we expect the field of diamond-based quantum optics to be an immediate beneficiary of this work, with the optical-microwave photon interface being key to the initialization, manipulation and detection of quantum states.

Online Content Methods, along with any additional Extended Data display items and Source Data, are available in the online version of the paper; references unique to these sections appear only in the online paper.

Received 26 October 2017; accepted 19 January 2018.

- Oxborrow, M., Breeze, J. & Alford, N. Room-temperature solid-state maser. *Nature* **488**, 353–356 (2012).
- Breeze, J. *et al.* Enhanced magnetic Purcell effect in room-temperature masers. *Nat. Commun.* **6**, 6215 (2015).
- Salvadori, E. *et al.* Nanosecond time-resolved characterization of a pentacene-based room-temperature maser. *Sci. Rep.* **7**, 41836 (2017).
- Poklonski, N., Lapchuk, N. & Lapchuk, T. Inverted EPR signal from nitrogen defects in a synthetic diamond single crystal at room temperature. *JETP Lett.* **80**, 748–751 (2004).
- Poklonski, N. *et al.* Nitrogen-doped chemical vapour deposited diamond: a new material for room-temperature solid state maser. *Chin. Phys. Lett.* **24**, 2088–2090 (2007).
- Jin, L. *et al.* Proposal for a room-temperature diamond maser. *Nat. Commun.* **6**, 8251 (2015).
- Kraus, H. *et al.* Room-temperature quantum microwave emitters based on spin defects in silicon carbide. *Nat. Phys.* **10**, 157–162 (2014).
- Smith, W., Sorokin, P., Gelles, I. & Lasher, G. Electron-spin resonance of nitrogen donors in diamond. *Phys. Rev.* **115**, 1546–1552 (1959).
- Sorokin, P., Lasher, G. & Gelles, I. Cross relaxation studies in diamond. *Phys. Rev.* **118**, 939–945 (1960).
- Siegman, A. *Microwave Solid-State Masers* Ch. 4 (McGraw-Hill, 1964).
- Loubser, J. & van Wyk, J. Electron spin resonance in the study of diamond. *Rep. Prog. Phys.* **41**, 1201–1248 (1978).
- Gruber, A. *et al.* Scanning confocal optical microscopy and magnetic resonance on single defect centers. *Science* **276**, 2012–2014 (1997).
- Taylor, J. *et al.* High-sensitivity diamond magnetometer with nanoscale resolution. *Nat. Phys.* **4**, 810–816 (2008); erratum **7**, 270 (2011).
- Maze, J. *et al.* Nanoscale magnetic sensing with an individual electronic spin in diamond. *Nature* **455**, 644–647 (2008).
- Childress, L. *et al.* Coherent dynamics of coupled electron and nuclear spin qubits in diamond. *Science* **314**, 281–285 (2006).
- Takahashi, S., Hanson, R., van Tol, J., Sherwin, M. S. & Awschalom, D. D. Quenching spin decoherence in diamond through spin bath polarization. *Phys. Rev. Lett.* **101**, 047601 (2008).
- Jarmola, A., Acosta, V., Jensen, K., Chemerisov, S. & Budker, D. Temperature- and magnetic-field-dependent longitudinal spin relaxation in nitrogen-vacancy ensembles in diamond. *Phys. Rev. Lett.* **108**, 197601 (2012).
- Robledo, L., Bernien, H., van der Sar, T. & Hanson, R. Spin dynamics in the optical cycle of single nitrogen-vacancy centres in diamond. *New J. Phys.* **13**, 025013 (2011).
- Doherty, M. W. *et al.* The nitrogen-vacancy colour centre in diamond. *Phys. Rep.* **528**, 1–45 (2013).
- Breeze, J. D. *et al.* Room-temperature cavity quantum electrodynamics with strongly coupled Dicke states. *npj Quantum Inf.* **3**, 1 (2017).
- Schawlow, A. L. & Townes, C. H. Infrared and optical masers. *Phys. Rev.* **112**, 1940–1949 (1958).
- Dimer, F., Estienne, B., Parkins, A. & Carmichael, H. Proposed realization of the Dicke-model quantum phase transition in an optical cavity QED system. *Phys. Rev. A* **75**, 013804 (2007).

Acknowledgements We thank J. Hall and M. Markham (Element 6 Ltd) for supplying the diamond samples, P. French and R. Taylor (Photonics Group at Imperial College London) for lending us their continuous-wave laser, and E. Bauch (Harvard University) for discussions. We also thank M. Lennon (IC), D. Halpin and D. Farquharson (UCL) for manufacturing the cavity components. This work was supported by the UK Engineering and Physical Sciences Research Council through grants EP/K011987/1 (IC) and EP/K011804/1 (UCL). We also acknowledge support from the Henry Royce Institute.

Author Contributions J.D.B. conceived the study, developed the theory, designed the maser cavity, devised the experiment and wrote software for collecting experimental data. J.D.B. and C.W.M.K. developed the experimental design and performed experiments with input from E.S. and J.S. J.D.B. interpreted the results with input from E.S. and C.W.M.K. J.S. characterized the diamond NV concentration by optical means and developed the optical pumping scheme. J.D.B., E.S. and C.W.M.K. characterized the diamonds using EPR. J.D.B. wrote the paper with assistance from C.W.M.K. and with additional editing by E.S. and N.M.A.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Publisher's note: Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations. Correspondence and requests for materials should be addressed to J.D.B. (jonathan.breeze@imperial.ac.uk).

Reviewer Information *Nature* thanks A. Blank, F. Jelezko and R.-B. Liu for their contribution to the peer review of this work.

METHODS

Spin Hamiltonian. The spin Hamiltonian for NV[−] centres in diamond is given by²³

$$H_{\text{spin}} = \gamma_e \mathbf{S} \cdot \mathbf{B} - \gamma_n \mathbf{I} \cdot \mathbf{B} + D \left[S_z^2 - \frac{1}{3} S(S+1) \right] + \mathbf{S} \mathbf{A} \mathbf{I}$$

where γ_e and γ_n are the gyromagnetic ratios of electrons and nuclei, respectively, $D \approx 2.87$ GHz is the fine structure zero-field splitting, \mathbf{S} and \mathbf{I} are the triplet spin and ¹⁴N nuclear spin eigenvectors, respectively, \mathbf{A} is the uniaxially anisotropic hyperfine coupling tensor ($A_{\perp} = -2.7$ MHz, $A_{\parallel} = -2.1$ MHz with respect to the NV axis) and \mathbf{B} is the magnetic field.

Spin-photon maser dynamics. The maser is modelled as an ensemble of N two-level quantum emitters (lower state $|g\rangle \equiv |-1\rangle$ and upper state $|e\rangle \equiv |0\rangle$) resonantly coupled to a cavity mode. The distance between the emitters is much less than the wavelength of the cavity mode and their spin-photon coupling is assumed to be homogeneous. This interaction can be described by the Tavis–Cummings Hamiltonian²⁴ within the rotating-wave approximation:

$$H_{\text{TC}} = \hbar\omega_c a^\dagger a + \frac{1}{2} \hbar\omega_s S^z + \hbar g_e (S^+ a + a^\dagger S^-)$$

where $g_e = g_s \sqrt{N}$ is the enhanced collective spin–photon coupling strength, g_s is the single spin–photon coupling strength, ω_c is the cavity frequency, ω_s is the (two-level) transition frequency of the spin, a^\dagger (a) is the creation (annihilation) operator of the cavity photon, S^z is the collective inversion operator, S^\pm are the normalized collective spin operators and \hbar is the reduced Planck constant. The single spin–photon coupling can be written as²⁵

$$g_s = \gamma_e \sqrt{\frac{\mu_0 \hbar \omega_c \kappa_s / (\kappa_s + \kappa_c)}{2V_m (1 + \Delta^2)}}$$

where $\Delta = 2(\omega_c - \omega_s)/(\kappa_s + \kappa_c)$ is a spin–cavity detuning parameter. The Hamiltonian, coupled with a dissipative Liouvillian, permits Lindblad master equations to be developed for the expectation values of the cavity field operator a , inversion S^z and transverse spin S^\pm operators. These can be solved in the steady state, leading to a simple maser threshold pumping rate per NV centre of

$$w_{\text{thr}} = \gamma(C - 1)^{-1}\eta$$

where $C = 4g_s^2 N / (\kappa_s \kappa_c)$ is the cooperativity, $\gamma = 1/T_1$ is the spin–lattice relaxation rate and η is a scaling factor based on the number of photons required to increase the inversion by one. The scaling factor η can be derived from the steady-state optical pump rate equations. We derived a value of $\eta \approx 14.4$; it holds for pump rates up to w_{thr} at which the populations of the $|\pm 1\rangle$ states are approximately equal. This is the maser amplification regime in which the inversion, given by $S^z = wN/(w + \eta\gamma)$, is positive. Accessing this regime is not challenging because the dark-state populations of the triplet ground states have approximately equal Boltzmann populations, which require only weak optical pumping for the inversion to become positive. One noticeable aspect of the threshold equation is that w_{thr} is lower than the scaled spin–lattice relaxation rate $\gamma\eta$ if $C > 1$, which can be considered the condition for masing. The inversion in the maser oscillation regime

$$S^z = N_e - N_g = \kappa_s \kappa_c / (4g_s^2) = N/C$$

is always less than N when $C > 1$. Above threshold and for steady-state conditions, we solve a nonlinear system of rate equations comprising seven NV[−] states (triplet ground, triplet excited and singlet), a cavity field operator \hat{a} , a longitudinal (inversion) operator \hat{S}^z and a transverse spin operator \hat{S}^\pm . These equations are the Maxwell–Bloch equations coupled to a set of optical spin dynamical rate equations. **Diamond preparation and characterization.** A synthetic diamond with natural carbon isotopic abundance, doped with approximately 5 p.p.m. of nitrogen donors was grown by chemical vapour deposition. The diamond was irradiated by 4.5-MeV electrons, equivalent to a dose of $5 \times 10^{18} \text{ cm}^{-2}$, to create vacancies. It was then annealed for 2 h at 400 °C, 16 h at 800 °C and a further 2 h at 1,200 °C to promote the formation of NV defect centres. The diamond was laser-cut into a cuboid of dimensions 2.1 mm \times 2.1 mm \times 2.6 mm, with {100} faces and <111> crystal directions (along which the NV centres lie) emerging from the corners of the cuboid. These steps were performed by Element 6 (UK). The diamond was cleaned in nitric acid to remove residual graphitic carbon (48 h) and then polished using a lint-free cloth.

NV defect centres exist in neutral NV⁰ and charged NV[−] states. Using ultraviolet–visible absorbance spectroscopy (Cary 5000 UV–VIS–NIR spectrophotometer, Ocean Optics USB 2000+ fibre optic spectrometer), the NV[−] concentration in the

sample was found to be 0.36 p.p.m. Further photoluminescence studies revealed that the ratio of charged to neutral NV centres (NV[−]/NV⁰) was approximately 0.55 under illumination from a 100-mW continuous-wave 532-nm Nd:YAG laser. Furthermore, the ratio decreased slightly as the laser power was reduced, suggesting that some NV[−] states had been photo-converted into NV⁰ ones^{26,27}. The total concentration of NV⁰ and NV[−] states was $[\text{NV}] \approx 1$ p.p.m., which, given the initial nitrogen concentration of $[\text{N}^0] \approx 5$ p.p.m., equates to a N-to-NV conversion efficiency of 20%. The final nitrogen (P1 centre) concentration would be $[\text{N}^0] \approx 4$ p.p.m. in this case.

EPR spectroscopy of NV centres. EPR spectroscopy at 9.5 GHz was used to align the diamond sample and measure the relaxation times T_1 , T_2 (Bruker Elexsys E580 spectrometer with ER 4118 X-MD5 resonator) and T_2^* (Bruker EMXplus with ER 4122SHQ resonator) of the high-field line. The diamond was held inside a hollow low-loss quartz tube and oriented inside the cylindrical single-crystal sapphire ring of the resonator so that two NV axes lay in a plane perpendicular to the cylindrical axis. The diamond was rotated about the cylindrical axis of the cavity until the magnetic field was aligned with one of the two <111> directions and colinear with a NV-defect axis. This was achieved by sweeping the magnetic field from 200 mT to 450 mT and noting the minimum and maximum fields at which absorption and emission occurred. When the difference in magnetic field between the low-field and high-field lines was equal to twice the zero-field splitting $4\pi D/\gamma_e$ or 205 mT, the absorption peaks associated with the other three <111> directions coalesced, becoming quasi-degenerate at about 370 mT. Pronounced low-field absorption and high-field emission peaks appeared at 239 mT and 444 mT, respectively. Nitrogen spin-1/2 (P1) impurities were evident at 341 mT. The spin–lattice relaxation time (T_1) was measured using an inversion recovery sequence π – T – $\pi/2$ – τ – π – τ –echo, in which the interpulse delay τ was kept fixed at 1.2 μ s and the time interval T was incremented in steps of 10 μ s from an initial value of 1 μ s. The length of the $\pi/2$ pulse was 200 ns and each dataset comprised 1,024 points. The magnetic field was set on resonance with the $m_I = 0$ hyperfine line of the highest EPR band (Fig. 1c). The homogeneous spin relaxation time (T_2) was measured using a Hahn-echo sequence $\pi/2$ – τ – π – τ –echo, in which the interpulse delay τ was incremented in steps of 80 ns from an initial value of 1.2 μ s. These measurements were repeated for increasing levels of optical pump power (Extended Data Fig. 2b, c). The inhomogeneous spin dephasing time T_2^* was measured using the saturation broadening technique (Extended Data Fig. 2d). The $m_I = 0$ linewidth was measured as a function of applied microwave power. By extrapolating a linear fit of linewidth-squared versus microwave power down to zero power, the inhomogeneous broadening and hence T_2^* can be found.

Maser cavity. A high-Purcell-factor cavity was designed using a radial mode-matching technique²⁸. The cavity comprised a single-crystal sapphire dielectric ring resonator (relative electrical permittivity $\epsilon_r = 9.394$; outer diameter, 10 mm; inner diameter, 5.1 mm; height, 6.0 mm) housed centrally within an oxygen-free high-conductivity copper cavity (diameter, 36 mm; height, 28 mm). The cavity supported a TE₀₁₈ mode resonating at 9.22 GHz with an unloaded Q factor of 55,000. The central bore containing the diamond had an electric filling factor of 0.008, allowing the dielectric properties of the diamond to be neglected. A magnetic mode volume of $V_m = 0.15 \text{ cm}^3$ was calculated from the ratio of the stored magnetic energy within the cavity to the maximum magnetic field energy density:

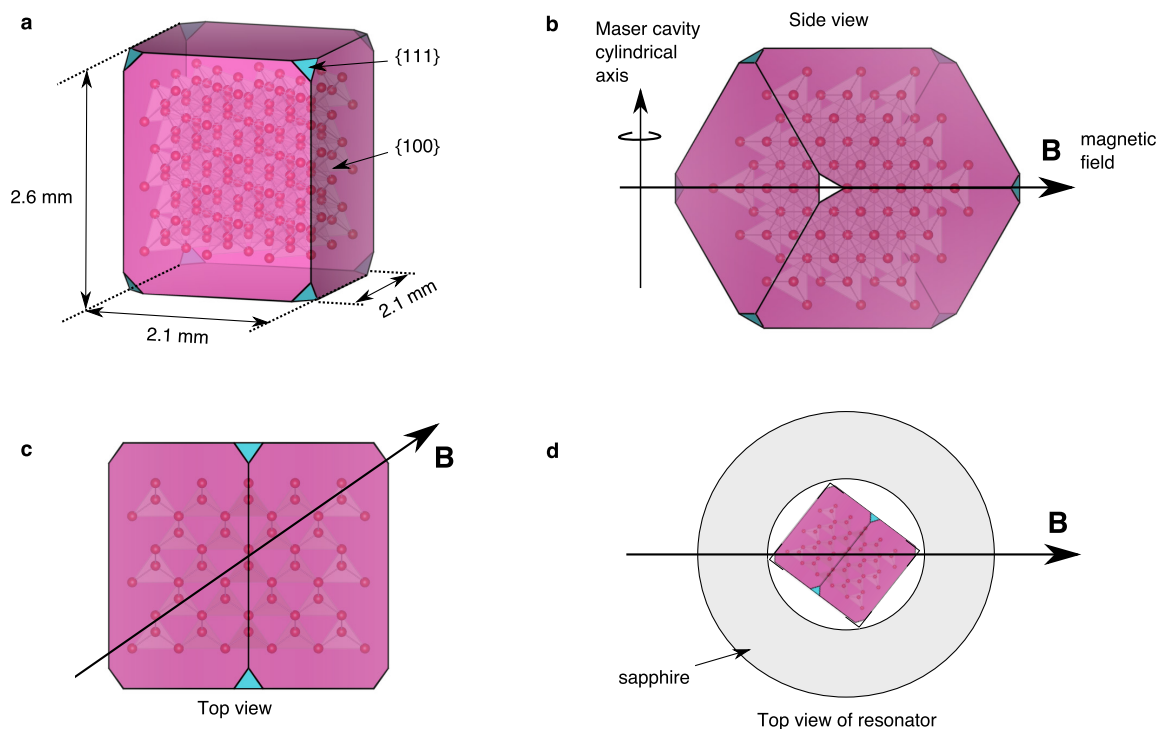
$$V_m = \int_V \frac{|\mathbf{H}(\mathbf{r})|^2}{|\mathbf{H}_{\text{max}}(\mathbf{r})|^2} dV$$

The electric filling factor of the sapphire was 0.76 and the magnetic filling factor of the central hole was 0.14. The diamond was mounted inside the sapphire resonator and held in place by quartz tubes. Coupling to the cavity was achieved by a small adjustable antenna loop, set slightly below critical coupling ($k = 0.85$) to yield a loaded Q factor of 30,000.

Optical pumping. A continuous-wave 532-nm laser was used to pump the NV[−] centres (Laser Quantum Finesse Pure) with a spot size of 2.5 mm. The laser beam was adjusted until it was colinear with the cavity axis and p -polarized with respect to the surface of the diamond, because the angle of incidence was close to the Brewster angle (67.2°). The one-photon absorption cross-section of NV[−] centres at 532 nm is $\sigma = 3.1 \times 10^{-17} \text{ cm}^2$ (ref. 29). The power required to pump the NV centres at rate w is therefore $P_p \approx \hbar w \omega_p (A/\sigma)$, where ω_p is the pump frequency and A is the area (spot size). The defect pump rate $w = \sigma I / (\hbar \omega_p)$, where ω_p the optical pump frequency, is proportional to the local pump intensity $I = P_p/A$, which varies across the sample.

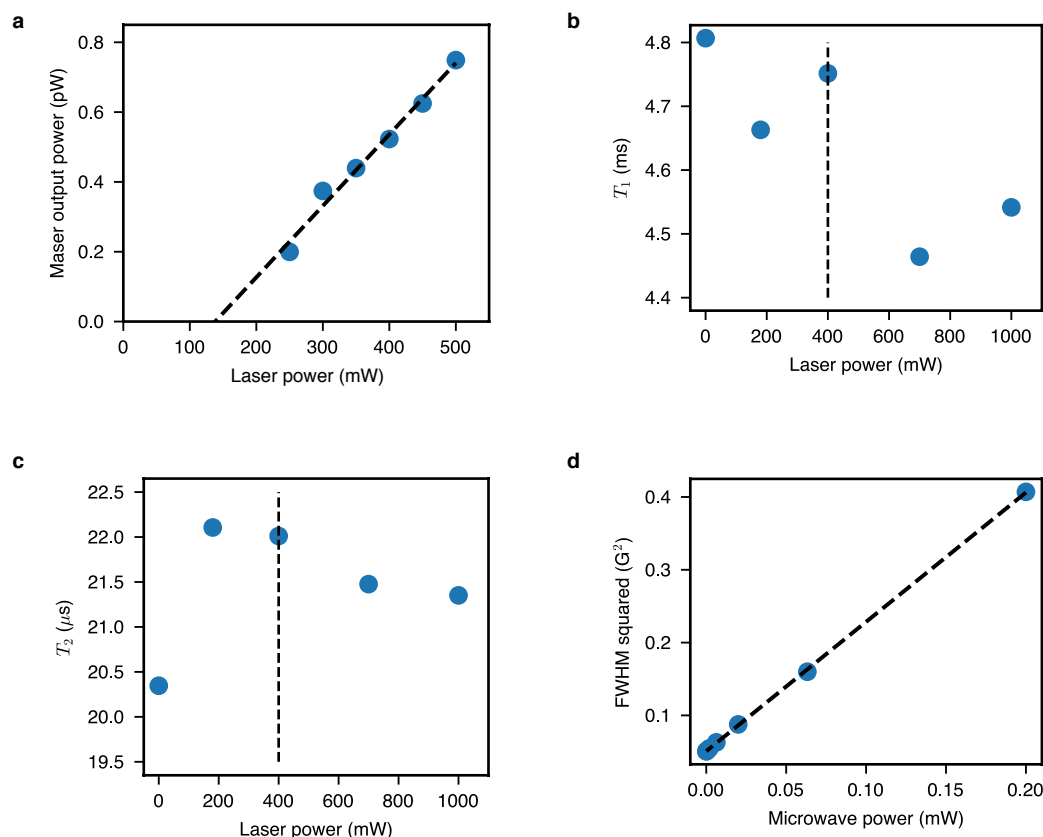
Data availability. The data that support the findings of this study are available from the corresponding author on reasonable request.

23. Yavkin, B., Mamin, G. & Orlinskii, S. High-frequency pulsed endor spectroscopy of the NV⁻ centre in the commercial HPHT diamond. *J. Magn. Reson.* **262**, 15–19 (2016).
24. Tavis, M. & Cummings, F. W. Exact solution for an *N*-molecule radiation-field Hamiltonian. *Phys. Rev.* **170**, 379–384 (1968).
25. Carmichael, H. J. *Statistical Methods in Quantum Optics 1: Master Equations and Fokker-Planck Equations* Ch. 7 (Springer, 2003).
26. Manson, N. & Harrison, J. Photo-ionization of the nitrogen-vacancy center in diamond. *Diamond Related Materials* **14**, 1705–1710 (2005).
27. Aslam, N., Waldherr, G., Neumann, P., Jelezko, F. & Wrachtrup, J. Photo-induced ionization dynamics of the nitrogen vacancy defect in diamond investigated by single-shot charge state detection. *New J. Phys.* **15**, 013064 (2013).
28. Kajfez, D. & Guillon, P. *Dielectric Resonators* 1st edn, Ch. 5 (Artech House, 1964).
29. Wee, T.-L. *et al.* Two-photon excited fluorescence of nitrogen-vacancy centers in proton-irradiated type Ib diamond. *J. Phys. Chem. A* **111**, 9379–9386 (2007).



Extended Data Figure 1 | Diamond sample geometry and orientation. **a**, The diamond is a rectangular cuboid with dimensions $2.1 \text{ mm} \times 2.1 \text{ mm} \times 2.6 \text{ mm}$. The main faces (pink) are $\{100\}$, with the $\{111\}$ faces (to which the NV $\langle 111 \rangle$ directions are normal) depicted in blue. **b**, Side view of the diamond oriented within the maser cavity, with static magnetic field B applied across the NV $\langle 111 \rangle$ direction. The microwave

magnetic field is perpendicular to the N–V defect axis and static magnetic field direction. **c**, Top view of the diamond when it is oriented inside the cavity. **d**, Top view of the diamond placed within the single-crystal sapphire ring within the maser cavity. The depiction of the structure of diamond within the crystals (carbon atoms, bonds and tetrahedral) are for illustrative purposes only.



Extended Data Figure 2 | Threshold and spin-relaxation measurements.

a, Maser threshold. The peak maser output power increases linearly as a function of the optical pump power (data). Extrapolation of the linear fit (dashed line) to zero maser output power reveals a threshold optical pump power of 138 mW, which is lower than the predicted 180 mW. **b**, Spin-lattice relaxation time T_1 as a function of laser pump power (data). The slight decrease in T_1 is expected and due to an increase in temperature caused by the non-radiative (heating) processes during the NV spin-polarizing optical pump cycle. **c**, Spin-decoherence time T_2 as a function of laser pump power (data). There is little change in T_2 , with a slight jump upon applying optical pumping, probably due to an increase in EPR signal

amplitude and hence less error. There is subsequently a slight decrease due to temperature increase and pumping decoherence. Photo-conversion of NV^- to NV^0 could also be a source of decoherence. **d**, Power saturation broadening. The inhomogeneous spin decoherence time $T_2^* = 2/\kappa_s$ was inferred from power saturation broadening measurements of the spin resonance lines. The spectral full-width at half-maximum (FWHM) γ was measured as a function of interrogating microwave power (data). A spin decoherence rate of $T_2^* = 0.52 \mu s$ was extracted by extrapolating the square of the FWHM linewidth down to zero microwave power (dashed line). The vertical dashed line in **b** and **c** depicts the applied optical pump power of 400 mW.

Maximizing and stabilizing luminescence from halide perovskites with potassium passivation

Mojtaba Abdi-Jalebi¹, Zahra Andaji-Garmaroudi¹, Stefania Cacovich², Camille Stavrakas¹, Bertrand Philippe³, Johannes M. Richter¹, Mejd Alsari¹, Edward P. Booker¹, Eline M. Hutter⁴, Andrew J. Pearson¹, Samuele Lilliu^{5,6}, Tom J. Savenije⁴, Håkan Rensmo³, Giorgio Divitini², Caterina Ducati², Richard H. Friend¹ & Samuel D. Stranks¹

Metal halide perovskites are of great interest for various high-performance optoelectronic applications¹. The ability to tune the perovskite bandgap continuously by modifying the chemical composition opens up applications for perovskites as coloured emitters, in building-integrated photovoltaics, and as components of tandem photovoltaics to increase the power conversion efficiency^{2–4}. Nevertheless, performance is limited by non-radiative losses, with luminescence yields in state-of-the-art perovskite solar cells still far from 100 per cent under standard solar illumination conditions^{5–7}. Furthermore, in mixed halide perovskite systems designed for continuous bandgap tunability² (bandgaps of approximately 1.7 to 1.9 electronvolts), photoinduced ion segregation leads to bandgap instabilities^{8,9}. Here we demonstrate substantial mitigation of both non-radiative losses and photoinduced ion migration in perovskite films and interfaces by decorating the surfaces and grain boundaries with passivating potassium halide layers. We demonstrate external photoluminescence quantum yields of 66 per cent, which translate to internal yields that exceed 95 per cent. The high luminescence yields are achieved while maintaining high mobilities of more than 40 square centimetres per volt per second, providing the elusive combination of both high luminescence and excellent charge transport¹⁰. When interfaced with electrodes in a solar cell device stack, the external luminescence yield—a quantity that must be maximized to obtain high efficiency—remains as high as 15 per cent, indicating very clean interfaces. We also demonstrate the inhibition of transient photoinduced ion-migration processes across a wide range of mixed halide perovskite bandgaps in materials that exhibit

bandgap instabilities when unpassivated. We validate these results in fully operating solar cells. Our work represents an important advance in the construction of tunable metal halide perovskite films and interfaces that can approach the efficiency limits in tandem solar cells, coloured-light-emitting diodes and other optoelectronic applications.

We fabricated a series of passivated triple-cation perovskite thin films on glass¹¹ ((Cs_{0.06}FA_{0.79}MA_{0.15})Pb(I_{0.85}Br_{0.15})₃, where MA = methylammonium, CH₃NH₃⁺ and FA = formamidinium, CH₃(NH₂)₂⁺), by diluting the precursor solution with potassium iodide solution. We denote the perovskite as (Cs,FA,MA)Pb(I_{0.85}Br_{0.15})₃ and the passivated samples as $x = [K]/([A] + [K])$, where A = (Cs, FA, MA); x represents the fraction of potassium ions out of the total monovalent cations in the precursor solution. We note that the standard triple-cation precursor solution recipe ($x = 0$) has a slight halide deficiency; however, the introduction of potassium iodide leads to samples with a slight excess of halide, along with very small changes to the I/Br ratio (Extended Data Fig. 1). The films have uniformly packed grains, each with a size of around 200–400 nm (Extended Data Fig. 1). Absorption and photoluminescence measurements reveal a reduction in the optical bandgap of the perovskite film upon increasing addition of potassium iodide, consistent with the selective interaction of the additives with bromide (Extended Data Fig. 2).

For a solar cell or light-emitting diode to approach its efficiency limit, all recombination should be radiative and the luminescence should be maximized¹². In state-of-the-art perovskite films, there are still substantial non-radiative losses that originate from charge-carrier trap

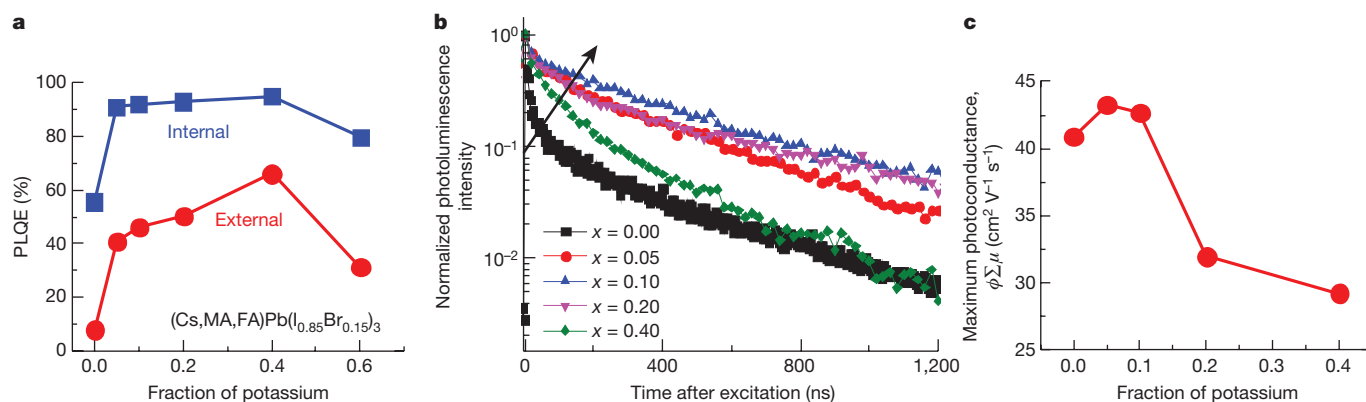


Figure 1 | Increased radiative efficiency and charge-carrier mobility upon passivation. **a**, PLQE of passivated perovskite thin films with increasing fraction of potassium x , measured under illumination with a 532-nm laser at an excitation intensity equivalent to approximately 1 sun (60 mW cm⁻²) after 300 s of illumination. **b**, Time-resolved

photoluminescence decays of the passivated perovskite films, with excitation at 400 nm and a pulse fluence of 0.5 μJ cm⁻² (excitation density of approximately 10¹⁶ cm⁻³). **c**, Maximum photoconductance for each of the potassium contents, extracted from TRMC measurements with an excitation density of approximately 10¹⁴ cm⁻³ (Extended Data Fig. 7).

¹Cavendish Laboratory, Department of Physics, University of Cambridge, JJ Thomson Avenue, Cambridge CB3 0HE, UK. ²Department of Materials Science and Metallurgy, University of Cambridge, 27 Charles Babbage Road, Cambridge CB3 0FS, UK. ³Department of Physics and Astronomy, Uppsala University, Box 516, 75120 Uppsala, Sweden. ⁴Department of Chemical Engineering, Delft University of Technology, van der Maasweg 9, 2629 HZ Delft, The Netherlands. ⁵Department of Physics and Astronomy, University of Sheffield, Sheffield S3 7RH, UK. ⁶The UAE Centre for Crystallography, United Arab Emirates.

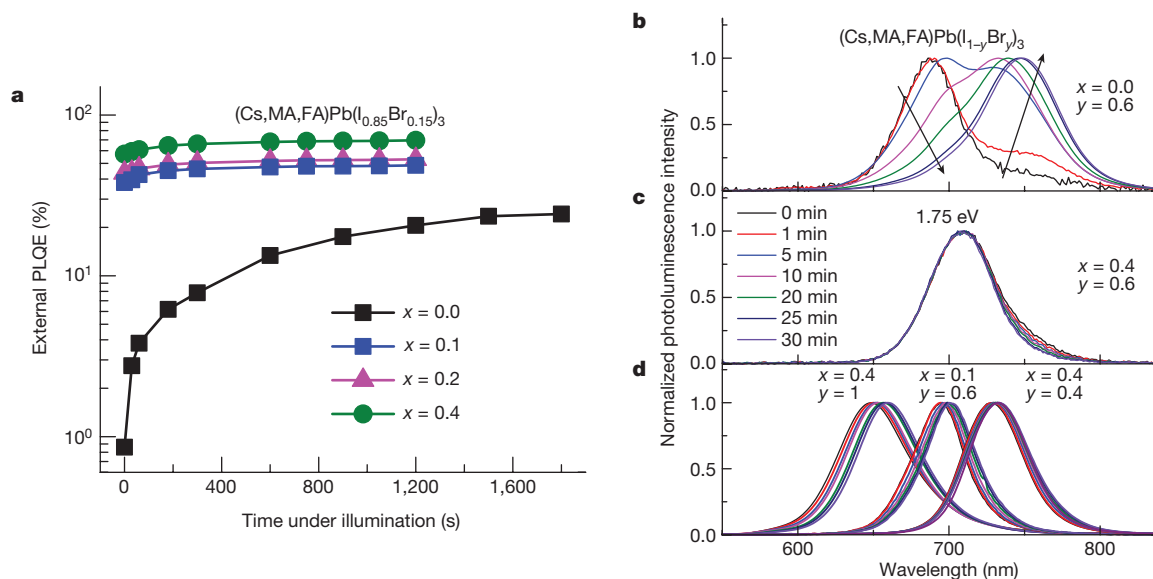


Figure 2 | Stabilized PLQE and inhibition of photoinduced ion migration. **a**, PLQE time course for $(\text{Cs,FA,MA})\text{Pb}(\text{I}_{0.85}\text{Br}_{0.15})_3$ films illuminated with a 532-nm laser at an excitation intensity equivalent to approximately 1 sun (60 mW cm^{-2}) in an ambient atmosphere. **b, c**, Photoluminescence from $(\text{Cs,FA,MA})\text{Pb}(\text{I}_{1-y}\text{Br}_y)_3$, with $y = 0.6$,

states present in the perovskite layer¹³. The origin of the trap states is unclear, but they may be associated with ionic defects such as halide vacancies^{14,15}. In Fig. 1a, we show the external photoluminescence quantum efficiency (PLQE) of the $(\text{Cs,FA,MA})\text{Pb}(\text{I}_{0.85}\text{Br}_{0.15})_3$ perovskite films with increasing potassium content measured at excitation densities equivalent to that of solar illumination. The PLQE shows a substantial increase from 8% when $x = 0$ to 41% when $x = 0.05$, reaching a very high PLQE of 66% when $x = 0.40$. By accounting for photon recycling and light-out-coupling effects¹⁶, these values translate to an internal PLQE that exceeds 95% for the passivated compositions (Fig. 1a). Furthermore, the PLQE does not vary considerably with excitation power for the passivated samples, unlike for the $x = 0$ sample

without passivation ($x = 0$, **b**) and with passivation ($x = 0.4$, **c**), illuminated continuously under the same conditions as in **a**. **d**, The photoluminescence from the passivated ($x = 0.4$) compositions with $y = 0.4$, $x = 0.4$ (peak at 1.70 eV), $y = 0.6$, $x = 0.1$ (1.78 eV) and $y = 1$, $x = 0.4$ (1.89 eV), measured over time under the same conditions.

in which the PLQE increases with increasing excitation power owing to filling of the trap states¹⁷ (Extended Data Fig. 2). These results are also observed in micro-photoluminescence measurements (Extended Data Fig. 3). Time-resolved photoluminescence experiments (Fig. 1b) show the removal of the fast non-radiative decay component with passivation, leading to radiative bimolecular recombination (Extended Data Fig. 4).

We used time-resolved microwave conductivity (TRMC) to assess the impact of passivation on charge transport in the $(\text{Cs,FA,MA})\text{Pb}(\text{I}_{0.85}\text{Br}_{0.15})_3$ perovskite thin films^{10,18} (Extended Data Fig. 5). In Fig. 1c, we show the maximum photoconductance (charge mobility) for each of the fractions of potassium. For $x = 0$ and $x = 0.1$, we find that the carrier mobility remains approximately constant at a high value of around

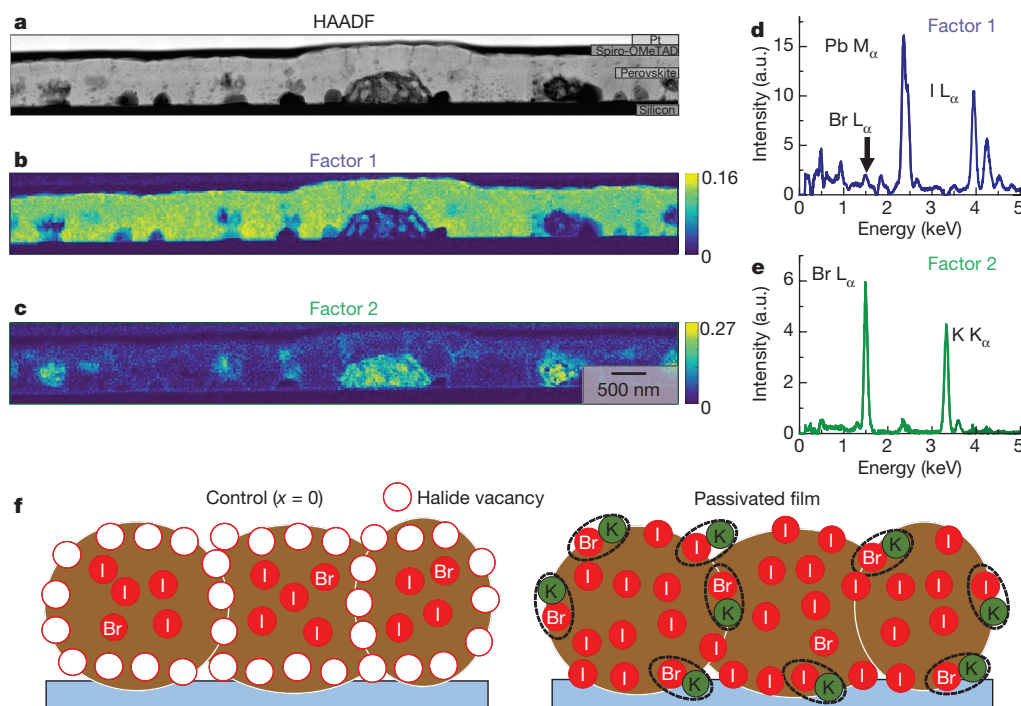


Figure 3 | Cross-sectional chemical characterization of a passivated perovskite thin film. **a**, HAADF STEM cross-sectional image of a $(\text{Cs,FA,MA})\text{Pb}(\text{I}_{0.85}\text{Br}_{0.15})_3$ passivated perovskite thin film ($x = 0.20$). **b, c**, Analysis of STEM data using NMF decomposition reveals the presence of factor 1 (b), associated with the perovskite layer, and of factor 2 (c), which is rich in potassium and bromide. **d, e**, The EDX spectra of factor 1 (d) and factor 2 (e). **f**, Schematic of a cross-section of a film showing halide-vacancy management in cases of excess halide, in which the surplus halide is immobilized through complexing with potassium into benign compounds at the grain boundaries and surfaces.

$42 \text{ cm}^2 \text{ V}^{-1} \text{ s}^{-1}$; this value decreases at higher potassium content, reaching around $30 \text{ cm}^2 \text{ V}^{-1} \text{ s}^{-1}$ when $x = 0.4$, which could be an effect of decreasing grain sizes when $x > 0.2$ (see Methods)¹⁹. These results are consistent with lower trap densities¹³ for the passivated samples, with the elimination of almost all non-radiative channels and the retention of excellent charge transport up to $x = 0.1$.

In Fig. 2a, we show the PLQE from the $(\text{Cs,FA,MA})\text{Pb}(\text{I}_{0.85}\text{Br}_{0.15})_3$ thin films as a function of time, under continuous illumination with intensity equivalent to 1 sun. For the reference film ($x = 0$), we observed a substantial but slow increase in PLQE with time, associated with photoinduced halide migration²⁰. By contrast, the high values of PLQE for the passivated films are stable under continuous illumination, suggesting that the photoinduced migration processes are substantially inhibited. To investigate the latter claim further, we prepared large-bandgap perovskite films by adding potassium iodide to precursor solutions with higher fractions of bromide; when unpassivated, such films typically show substantial photoluminescence shifts due to photoinduced halide segregation and subsequent emission from low-bandgap iodide-rich components⁸. We show that the photoluminescence spectral output of passivated $(\text{Cs,FA,MA})\text{Pb}(\text{I}_{0.4}\text{Br}_{0.6})_3$ films is very stable at the optimal bandgap for perovskite-silicon tandems² (1.75 eV) under 1-sun illumination (Fig. 2c). By contrast, the sample without passivation shows substantial redshifts and bandgap changes over time (Fig. 2b). In Fig. 2d, we show that this photostability is also seen across bromide fractions covering the range of idealized wide bandgaps for perovskite-perovskite tandems² (1.7–1.9 eV), albeit with slightly reduced stability at the highest bromide fractions (Extended Data Fig. 4). We also found that the critical bandgaps for tandems can be stabilized even at low passivation levels ($x = 0.1$, Fig. 2d). To our knowledge, this is the first report that shows such high stability in mixed halide compositions across a wide range of bandgaps under solar illumination in ambient conditions³.

Next, we investigated the chemical composition of the passivated $(\text{Cs,FA,MA})\text{Pb}(\text{I}_{0.85}\text{Br}_{0.15})_3$ perovskite thin films using scanning transmission electron microscopy–energy dispersive X-ray spectroscopy (STEM–EDX). In Fig. 3a, we show a cross-sectional view of a lamella of a film of composition $x = 0.2$. From the STEM–EDX elemental analysis, we observe a potassium-rich phase at the grain boundaries of the perovskite as well as at the interface with the substrate (Fig. 3c and

Table 1 | Parameters of the best-performing photovoltaic devices

| Potassium fraction, x | Bandgap (eV) | J_{sc} (mA cm^{-2}) | V_{oc} (V) | Fill factor | PCE (%) | V_{oc} loss (V) |
|-----------------------------------------------------------------------------------|--------------|----------------------------------|--------------|-------------|---------|-------------------|
| $(\text{Cs,MA,FA})\text{Pb}(\text{I}_{0.85}\text{Br}_{0.15})_3$ | | | | | | |
| 0.0 | 1.59 | 22.6 | 1.05 | 0.73 | 17.3 | 0.26 |
| 0.1 | 1.56 | 23.2 | 1.17 | 0.79 | 21.5 | 0.11 |
| $(\text{Cs,MA,FA})\text{Pb}(\text{I}_{0.4}\text{Br}_{0.6})_3$ | | | | | | |
| 0.0 | 1.83 | 15.3 | 1.12 | 0.72 | 12.3 | 0.42 |
| 0.1 | 1.78 | 17.9 | 1.23 | 0.79 | 17.5 | 0.27 |

The bandgaps are extracted from the onset of the external quantum efficiencies, and the V_{oc} loss is the difference between the bandgap-radiative-limit V_{oc} and the measured V_{oc} .

Extended Data Fig. 6). Analysis of the dataset using a non-negative matrix factorization (NMF) algorithm²¹ highlights the presence of two compositional phases in the specimen, denoted factor 1 and factor 2 in Fig. 3b, c. Factor 1 shows characteristic EDX features of the perovskite phase, including Br L_{α} , Pb M_{α} and I L_{α} lines (Fig. 3d), whereas factor 2 is rich in bromide and potassium (Fig. 3e) and, notably, is particularly prominent at the grain boundaries and at the top and bottom surfaces of the perovskite film. This is probably related to a new crystalline phase that is observed in grazing-incidence wide-angle X-ray scattering (GIWAXS) experiments (Extended Data Fig. 7). These results are also consistent with hard-X-ray photoelectron spectroscopy (HAXPES) measurements, which reveal a decrease in potassium upon moving from the film surface into the bulk (Extended Data Fig. 8). Collectively, these results indicate the formation of potassium halide (particularly bromide-rich) passivation layers decorating the surfaces, where the potassium is not incorporated into the perovskite lattice. This is in contrast to previous works^{5,22–25} that report the addition of small monovalent cations—including sodium, rubidium or potassium—to the perovskite; these studies proposed that the cations were incorporated into the lattice, although a recent report²⁶ shows that rubidium does not incorporate and suggests the same for potassium.

In Fig. 3f, we summarize our interpretation of the collective results. We used potassium iodide to introduce excess iodide into the perovskite precursor solutions, to compensate for any halide vacancies. The excess halides fill these vacancies, thereby passivating the non-radiative recombination pathways and leading to high luminescence

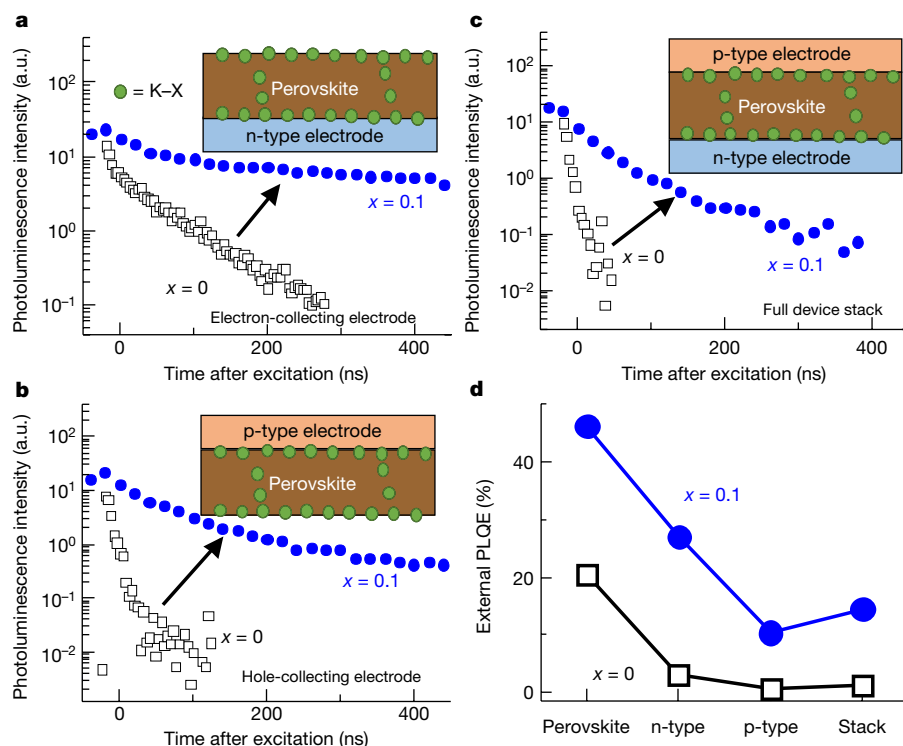


Figure 4 | Luminescence properties of the perovskite when interfaced with solar cell device contacts. a–c, Time-resolved photoluminescence decays of encapsulated $(\text{Cs,FA,MA})\text{Pb}(\text{I}_{0.85}\text{Br}_{0.15})_3$ films ($x = 0$ and $x = 0.1$) with excitation at 400 nm and a pulse fluence of $0.05 \mu\text{J cm}^{-2}$ ($1.5 \times 10^{15} \text{ cm}^{-3}$, equivalent to about 1 sun), when the perovskite is interfaced with an n-type electron-collecting electrode (compact- TiO_2 /thin-mesoporous TiO_2) (a), a p-type hole-collecting electrode (spiro-OMeTAD) (b) and both electrodes in a full device stack (c). d, External PLQE measurements of the perovskite in each configuration measured under illumination with a 532-nm laser at an excitation intensity equivalent to approximately 1 sun (60 mW cm^{-2}). See Extended Data Fig. 9 for other intensities.

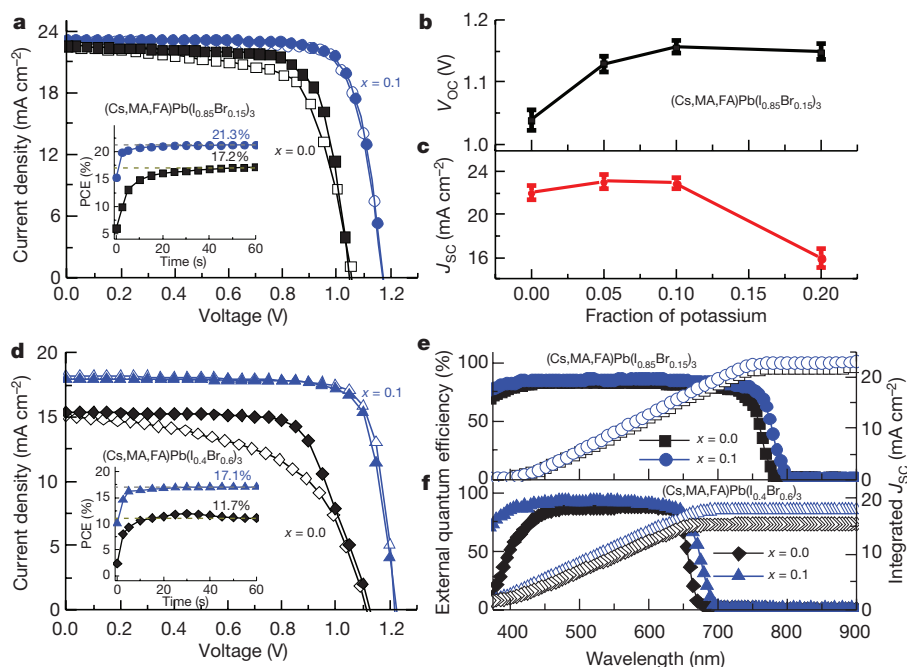


Figure 5 | Enhanced solar-cell power conversion efficiency. **a**, Forward (open symbols) and reverse (closed symbols) J - V curves of the best-performing solar cells with $(\text{Cs,MA,FA})\text{Pb}(\text{I}_{0.85}\text{Br}_{0.15})_3$ absorbers without ($x = 0$) and with ($x = 0.1$) passivation, measured under full simulated solar illumination conditions (AM1.5, 100 mW cm^{-2}). Inset, stabilized power output under the same conditions. **b**, **c**, Open-circuit voltage (V_{OC}) (**b**) and short-circuit current density (J_{SC}) (**c**) as functions of potassium fraction x , with error bars representing the standard deviation across ten devices for each composition. **d**, J - V curves of the best-performing solar cells with $(\text{Cs,MA,FA})\text{Pb}(\text{I}_{0.4}\text{Br}_{0.6})_3$ absorbers without ($x = 0$) and with ($x = 0.1$) potassium passivation. **e**, **f**, External quantum efficiencies and integrated short-circuit current for the $(\text{Cs,MA,FA})\text{Pb}(\text{I}_{0.85}\text{Br}_{0.15})_3$ (**e**) and $(\text{Cs,MA,FA})\text{Pb}(\text{I}_{0.4}\text{Br}_{0.6})_3$ (**f**) devices.

efficiencies. The excess halides are immobilized in the form of benign, potassium-rich, halide-sequestering species at the grain boundaries and surfaces, thereby inhibiting halide migration and suppressing any additional non-radiative decay arising from interstitial halides. At potassium contents greater than $x = 0.1$, these non-perovskite species perturb charge transport because of their large size (Extended Data Fig. 6). This suggests that there is an optimal potassium content of approximately 0.1, which is a compromise between high radiative efficiency and the retention of high charge-carrier mobility. Finally, we propose that potassium selectively depletes bromide from the perovskite crystal structure; this is consistent with the increased lattice parameter, redshifting band-edge, and decreasing Br/I ratio in the bulk observed upon the addition of potassium (Extended Data Figs 2 and 8). These observations and such substantially enhanced properties are not found in the absence of bromide (Extended Data Fig. 2). The addition of small fractions of bromide to the perovskite precursor solutions has been shown to improve perovskite film formation²⁷. However, bromide-rich perovskites typically have increased trap states and inferior charge-carrier mobility compared to their iodide-based counterparts²⁸. Here we exploit the beneficial properties of bromide in the grain-formation process, while suppressing the formation of bromide-induced defect states in the bulk of the crystal.

At open circuit in a solar cell, external luminescence should be maximized¹², and any additional non-radiative losses that occur upon the introduction of device electrodes must be minimized. In Fig. 4a, d, we show the time-resolved photoluminescence decays and PLQE, respectively, for the $(\text{Cs,FA,MA})\text{Pb}(\text{I}_{0.85}\text{Br}_{0.15})_3$ perovskite with and without potassium passivation when deposited on a standard n-type mesoporous TiO_2 electron-accepting contact (see Extended Data Fig. 9 for other excitation intensities). We found that charge-carrier recombination in the presence of the electrode is slower and more radiative (higher integrated time-resolved photoluminescence intensity) in the presence of potassium, with the PLQE decreasing by a factor of only 1.7 (to 27.1%), compared to a factor of 6.7 (to 3.0%) without passivation. In Fig. 4b, we show the time-resolved photoluminescence decays for a perovskite containing a top layer based on a standard p-type hole-accepting contact spiro-OMeTAD (2,2',7,7'-tetrakis(*N,N*-di-*p*-methoxyphenylamine)-9,9'-spirobifluorene). We again find that passivation leads to slower charge-carrier recombination and a less notable reduction in PLQE upon introduction of the electrode (Fig. 4d)—a decrease by a factor of 4.5 (to 10.4%) with passivation, compared to a decrease by a factor of 38 (to 0.5%) without. In Fig. 4c,

time-resolved photoluminescence decays for the full device stack (that is, both electrodes) clearly demonstrate that recombination is slower and the decrease in PLQE is mitigated upon passivation. The external PLQE of the full stack is 14.5% with the passivating interlayers (a reduction by a factor of 3.2 upon addition of the electrodes), which is an order of magnitude higher than the value of 1.2% for the control stack (reduction by a factor of 17). These results show that the potassium interlayers not only improve the optoelectronic properties of the perovskite material alone, but also lead to vastly improved interfaces with device electrodes.

To validate our findings, we constructed full solar cells using the device architecture fluorinated-tin oxide/compact- TiO_2 (approximately 30 nm)/thin-mesoporous TiO_2 (approximately 200 nm)/perovskite (approximately 500 nm)/spiro-OMeTAD (approximately 150 nm)/Au (80 nm). In Fig. 5a, we show the forward and reverse current-voltage (J - V) curves of the best-performing devices containing the $(\text{Cs,FA,MA})\text{Pb}(\text{I}_{0.85}\text{Br}_{0.15})_3$ absorbers with $x = 0$ and $x = 0.1$ under full simulated sunlight, with the extracted parameters given in Table 1 (see Extended Data Figs 10 and 11 and Table 1 for other potassium compositions, dark J - V curves and device statistics). We found that the device power conversion efficiency (PCE) increases from 17.3% ($x = 0$) to 21.5% ($x = 0.1$) upon passivation, with the elimination of hysteresis in the latter case consistent with inhibition of ion migration^{1,20}. This is also consistent with a rapid increase in stabilized power efficiency to a value of 21.3%, compared to a slower increase to just 17.2% for the control (Fig. 5a, inset). We observed an increase in open-circuit voltage (V_{OC}) upon passivation, from 1.05 V ($x = 0$) to 1.17 V ($x = 0.1$) (Fig. 5b); the calculated difference in V_{OC} (obtained from the increase in PLQE of the device stacks from 1.2% to 14.5%, Fig. 4d) matches within the error of the device statistics. The loss of open-circuit voltage from the radiative limit is only 0.11 V, and is one of the lowest losses reported in a perovskite solar cell so far⁵. We also see an increase in the short-circuit current density (J_{SC}) upon potassium addition up to a value of $x = 0.1$ (Fig. 5c), consistent with the increased carrier mobility and lifetime²⁹. The optimal device performance at $x = 0.1$ therefore validates the compromise between radiative efficiency and carrier mobility. We conducted preliminary stability tests and found a negligible drop in shelf-life performance over a month, and the devices retained over 80% of their initial performance after 300 h of continuous operation at maximum power (Extended Data Fig. 10). We show device results for larger-bandgap $(\text{Cs,FA,MA})\text{Pb}(\text{I}_{0.4}\text{Br}_{0.6})_3$ absorbers (Fig. 5d, f), attaining a PCE of 17.9% with minimal hysteresis for the $x = 0.1$ composition

and stabilized power output of 17.1% (see Table 1 and Extended Data Fig. 11 for device statistics). This is one of the highest device efficiencies yet reported for a large-bandgap (1.78 eV) perovskite that is ideally suited for tandem applications³.

These results are particularly notable for two reasons. First, the perovskite films and interfaces are surprisingly tolerant to passivating additives. Here, we have introduced additives at sufficiently high loading levels to passivate surfaces and stabilize luminescence across a range of bandgaps without compromising charge transport or extraction. This tolerance is in contrast to that of conventional semiconductors such as GaAs, which require more complicated passivation approaches such as controlled growth on lattice-matched substrates³⁰. Second, these results show directly the importance of obtaining high, stable external luminescence yields in full device stacks that contain luminescent absorbers capable of recycling photons^{12,31}. The internal luminescence yields approaching 100%, along with the small loss in external luminescence yield in the full device stack, shows that perovskites can sustain the necessary photon gas to achieve voltage losses low enough to rival those of GaAs. We note that the perovskite compositions and passivating interlayers used in this work lead to materials with far greater luminescence and broader stability enhancements compared with those containing rubidium or potassium that have been reported previously^{5,23–26}. Further investigation into potassium-based passivation will be required, including its efficacy in different perovskite compositions, deposition methods and contacts, and how it could be ultimately exploited (or even mimicked, using other halide-sequestering species) to eliminate interfacial non-radiative losses and ionic migration. The combination of high radiative efficiency, excellent charge transport and photostable bandgaps makes these passivation approaches extremely promising in taking perovskite devices to their efficiency limits across a range of bandgaps.

Online Content Methods, along with any additional Extended Data display items and Source Data, are available in the online version of the paper; references unique to these sections appear only in the online paper.

Received 28 October 2017; accepted 19 January 2018.

- Stranks, S. D. & Snaith, H. J. Metal-halide perovskites for photovoltaic and light-emitting devices. *Nat. Nanotechnol.* **10**, 391–402 (2015).
- Eperon, G. E. *et al.* Perovskite-perovskite tandem photovoltaics with optimized band gaps. *Science* **354**, 861–865 (2016).
- McMeekin, D. P. *et al.* A mixed-cation lead mixed-halide perovskite absorber for tandem solar cells. *Science* **351**, 151–155 (2016).
- Bush, K. A. *et al.* 23.6%-efficient monolithic perovskite/silicon tandem solar cells with improved stability. *Nat. Energy* **2**, 17009 (2017).
- Saliba, M. *et al.* Incorporation of rubidium cations into perovskite solar cells improves photovoltaic performance. *Science* **354**, 206–209 (2016).
- Momblona, C. *et al.* Efficient vacuum deposited p-i-n and n-i-p perovskite solar cells employing doped charge transport layers. *Energy Environ. Sci.* **9**, 3456–3463 (2016).
- Stranks, S. D. Nonradiative losses in metal halide perovskites. *ACS Energy Lett.* **2**, 1515–1525 (2017).
- Hoke, E. T. *et al.* Reversible photo-induced trap formation in mixed-halide hybrid perovskites for photovoltaics. *Chem. Sci.* **6**, 613–617 (2015).
- Barker, A. J. *et al.* Defect-assisted photoinduced halide segregation in mixed-halide perovskite thin films. *ACS Energy Lett.* **2**, 1416–1424 (2017).
- Stoddard, R. J., Eickemeyer, F. T., Katahara, J. K. & Hillhouse, H. W. Correlation between photoluminescence and carrier transport and a simple *in situ* passivation method for high-bandgap hybrid perovskites. *J. Phys. Chem. Lett.* **8**, 3289–3298 (2017).
- Saliba, M. *et al.* Cesium-containing triple cation perovskite solar cells: improved stability, reproducibility and high efficiency. *Energy Environ. Sci.* **9**, 1989–1997 (2016).
- Miller, O. D., Yablonovitch, E. & Kurtz, S. R. Strong internal and external luminescence as solar cells approach the Shockley–Queisser limit. *IEEE J. Photovoltaics* **2**, 303–311 (2012).
- Stranks, S. D. *et al.* Recombination kinetics in organic–inorganic perovskites: excitons, free charge, and subgap states. *Phys. Rev. Appl.* **2**, 034007 (2014).
- Kim, J., Lee, S. H., Lee, J. H. & Hong, K. H. The role of intrinsic defects in methylammonium lead iodide perovskite. *J. Phys. Chem. Lett.* **5**, 1312–1317 (2014).
- Agiorgousis, M. L., Sun, Y.-Y., Zeng, H. & Zhang, S. Strong covalency-induced recombination centers in perovskite solar cell material CH₃NH₃PbI₃. *J. Am. Chem. Soc.* **136**, 14570–14575 (2014).

- Richter, J. M. *et al.* Enhancing photoluminescence yields in lead halide perovskites by photon recycling and light out-coupling. *Nat. Commun.* **7**, 13941 (2016).
- Deschler, F. *et al.* High photoluminescence efficiency and optically pumped lasing in solution-processed mixed halide perovskite semiconductors. *J. Phys. Chem. Lett.* **5**, 1421–1426 (2014).
- Hutter, E. M. *et al.* Direct–indirect character of the bandgap in methylammonium lead iodide perovskite. *Nat. Mater.* **16**, 115–120 (2017).
- Reid, O. G., Yang, M., Kopidakis, N., Zhu, K. & Rumbles, G. Grain-size-limited mobility in methylammonium lead iodide perovskite thin films. *ACS Energy Lett.* **1**, 561–565 (2016).
- deQuilettes, D. W. *et al.* Photo-induced halide redistribution in organic–inorganic perovskite films. *Nat. Commun.* **7**, 11683 (2016).
- Pauca, V. P., Piper, J. & Plemmons, R. J. Nonnegative matrix factorization for spectral data analysis. *Linear Algebra Appl.* **416**, 29–47 (2006).
- Abdi-Jalebi, M. *et al.* Impact of monovalent cation halide additives on the structural and optoelectronic properties of CH₃NH₃PbI₃ perovskite. *Adv. Energy Mater.* **6**, 1502472 (2016).
- Tang, Z. *et al.* Hysteresis-free perovskite solar cells made of potassium-doped organometal halide perovskite. *Sci. Rep.* **7**, 12183 (2017).
- Zhao, P. *et al.* Improved carriers injection capacity in perovskite solar cells by introducing A-site interstitial defects. *J. Mater. Chem. A* **5**, 7905–7911 (2017).
- Nam, J. K. *et al.* Potassium incorporation for enhanced performance and stability of fully inorganic cesium lead halide perovskite solar cells. *Nano Lett.* **17**, 2028–2033 (2017).
- Kubicki, D. J. *et al.* Phase segregation in Cs-, Rb- and K-doped mixed-cation (MA)_x(FA)_{1-x}PbI₃ hybrid perovskites from solid-state NMR. *J. Am. Chem. Soc.* **139**, 14173–14180 (2017).
- Ibrahim Dar, M. *et al.* Understanding the impact of bromide on the photovoltaic performance of CH₃NH₃PbI₃ solar cells. *Adv. Mater.* **27**, 7221–7228 (2015).
- Sutter-Fella, C. M. *et al.* High photoluminescence quantum yield in band gap tunable bromide containing mixed halide perovskites. *Nano Lett.* **16**, 800–806 (2016).
- Stranks, S. D. *et al.* Electron–hole diffusion lengths exceeding 1 micrometer in an organometal trihalide perovskite absorber. *Science* **342**, 341–344 (2013).
- Yablonovitch, E. & Miller, O. D. The opto-electronics which broke the efficiency record in solar cells. In *Proc. Conference on Lasers and Electro-Optics 2012* CF2J.1 (Optical Society of America, 2012).
- Pazos-Outón, L. M. *et al.* Photon recycling in lead iodide perovskite solar cells. *Science* **351**, 1430–1433 (2016).

Acknowledgements M.A.-J. thanks Nava Technology Limited and Nyak Technology Limited for their funding and technical support. Z.A.-G. acknowledges funding from a Winton Studentship, and ICON Studentship from the Lloyd's Register Foundation. This project has received funding from the European Union's Seventh Framework Programme (FP7/2007–2013) under REA grant agreement number P10F-GA-2013-622630, the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation programme (grant agreement number 756962), and the Royal Society and Tata Group (UF150033). We thank the Engineering and Physical Sciences Research Council (EPSRC) for support. XMaS is a mid-range facility at the European Synchrotron Radiation Facility supported by the EPSRC and we are grateful to the XMaS beamline team staff for their support. We thank Diamond Light Source for access to beamline I09 and staff member T.-L. Lee as well as U. Cappel for assistance during the HAXPES measurements. S.C., C.D. and G.D. acknowledge funding from the ERC under grant number 25961976 PHOTO EM and financial support from the European Union under grant number 77 312483 ESTEEM2. M.A. thanks the president of the UAE's Distinguished Student Scholarship Program, granted by the Ministry of Presidential Affairs. H.R. and B.P. acknowledge support from the Swedish research council (2014–6019) and the Swedish foundation for strategic research. E.M.H. and T.J.S. were supported by the Netherlands Organization for Scientific Research under the Echo grant number 712.014.007.

Author Contributions M.A.-J. and S.D.S. conceived and planned the experiments with additional input from R.H.F. M.A.-J. fabricated all samples and devices and performed and analysed the PLQE experiments, the photoluminescence stability experiments, the time-resolved photoluminescence, absorption and photothermal deflection spectroscopies, the SEM and the device characterisation measurements. M.A.-J., Z.A.-G. and C.S. obtained and analysed the confocal photoluminescence maps. S.C., G.D. and C.D. performed and analysed the STEM–EDX measurements. M.A. and S.L. performed the GIWAXS experiments and analysed the data. E.M.H. and T.J.S. performed the TRMC measurements and analysed the data. M.A.-J., B.P. and H.R. performed and analysed HAXPES measurements. M.A.-J. and E.P.B. performed X-ray diffraction measurements and J.M.R. calculated the internal PLQE and assisted with time-resolved photoluminescence measurement and analysis. M.A.-J. and A.J.P. carried out the device stability tests. M.A.-J. and S.D.S. took the lead in drafting the manuscript and compiled the figures. All authors discussed the results and provided feedback on the manuscript.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Publisher's note: Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations. Correspondence and requests for materials should be addressed to S.D.S. (sds65@cam.ac.uk).

METHODS

Film and device fabrication. All the organic cation salts were purchased from Dyesol, the lead compounds were from TCI, and caesium iodide and potassium iodide were obtained from Alfa Aesar. spiro-OMeTAD was purchased from Borun Chemicals and was used as received. Unless otherwise stated, all other materials were purchased from Sigma-Aldrich.

The triple-cation-based perovskite $\text{Cs}_{0.06}\text{FA}_{0.79}\text{MA}_{0.15}\text{Pb}(\text{I}_{0.85}\text{Br}_{0.15})_3$ was prepared by dissolving PbI_2 (1.2 M), formamidinium iodide (1.11 M), methylammonium bromide (0.21 M) and PbBr_2 (0.21 M) in a mixture of anhydrous DMF:DMSO (4:1 volume ratio, v:v) followed by addition of 5 vol% from CsI stock solution (1.5 M in DMSO). The potassium iodide was first dissolved in a mixed solution of DMF:DMSO 4:1 (v:v) to make a stock solution of 1.5 M. We then added the potassium iodide solution into the triple cation perovskite solution in different volume ratios. We spin-coated the perovskite solutions using a two-step program at 2,000 and 6,000 rpm for 10 and 30 s, respectively, adding a 150 μl drop of chlorobenzene 30 s after the start of the spinning routine. We then annealed the films at 100 °C for 1 h. All the film preparations were performed in a nitrogen-filled glove box. The devices were fabricated following the same procedures for substrate preparation as well as deposition of both electron and hole transport layers (that is, TiO_2 , spiro-OMeTAD) as in ref. 32.

Scanning electron microscopy. The surface morphology of the films was examined using a field-emission scanning electron microscope (Merlin). An electron beam accelerated to 3 kV was used with an in-lens detector.

Steady-state absorption and photoluminescence characterization. Absorption spectra were recorded with a Perkin-Elmer Lambda 1050 spectrophotometer equipped with an integrating sphere to account for reflected and transmitted light. Photothermal deflection spectroscopy measurements were acquired on a custom-built set-up by monitoring the deflection of a fixed wavelength (670 nm) laser probe beam after absorption of each monochromatic pump wavelength by a thin film immersed in an inert liquid FC-72 Fluorinert (3M).

Photoluminescence quantum yield measurements were taken by mounting perovskite films or encapsulated device stacks in an integrating sphere and photo-exciting with a 532-nm continuous-wave laser. The laser and the emission signals were measured and quantified using a calibrated Andor iDUS DU490A InGaAs detector for the determination of photoluminescence quantum efficiency. The external PLQE was calculated as per ref. 33 and the internal PLQE was subsequently determined using the methods described in ref. 16.

The absorption and photoluminescence (Extended Data Fig. 2a, b) spectra reveal a reduction in the optical bandgap of the perovskite film upon the addition of increasing amounts of potassium iodide. The photoluminescence peaks at 770 nm for $x=0$ and redshifts to 807 nm for $x=0.4$. From HAXPES measurements, we find that increasing amounts of potassium iodide have no effect on the valence band edge, which suggests that it is the conduction band energy level that must become deeper (Extended Data Fig. 2b, inset). We note that this is in contrast to the case of the bandgap tunability achieved by directly changing mixed halide fractions, in which the valence band was shown to change but the conduction band remained almost fixed³⁴. Measurements on similar perovskite systems without bromide did not show such a strong redshift upon the addition of potassium (Extended Data Fig. 2f–i), which suggests that the additives are selectively interacting with the bromide. These results highlight the combination of potassium additives and mixed halides as levers for selectively tuning the bandgap and conduction band. We also note that the samples show a small fraction of material (less than 1%) with a low-bandgap component (Extended Data Fig. 2c), but the emission is predominantly from the slightly higher energy component. It is currently unclear why the luminescence is not dominated by the low-gap component, but it is possible that these low-concentration, low-bandgap components are electronically isolated from the remaining material, for example being surrounded by potassium-rich passivating material. We also note that the decrease in PLQE for $x=0.6$ could be due to the much smaller grain size at these high passivation levels (Extended Data Fig. 1). Finally, we note that changes in absorbance due to thickness changes upon potassium iodide addition are negligible over the ranges for our devices ($x=0$ –0.2), and therefore are not a major cause of the observed enhancements in device performance.

Confocal photoluminescence maps. Photoluminescence maps were collected using a WITec alpha 300 s setup. The excitation source was a 405-nm continuous-wave laser (Coherent CUBE), chopped using a Stanford Research SR 540 chopping unit at frequency of 840 Hz. The light was coupled through an optical fibre to the microscope and focused using a 100 \times Nikon lens (numerical aperture = 0.7). The sample was positioned on an X–Y piezo stage of the microscope. The photoluminescence signal was collected in reflection mode with the same 100 \times objective and detected using a spectrometer fitted with a CCD (charge-coupled device) detector. A low-pass filter with a cut-off wavelength of 435 nm was fitted before the CCD detector to block the excitation component of transmitted light (405 nm).

The confocal photoluminescence intensity maps are shown in Extended Data Fig. 3a–d from perovskite films with $x=0$ –0.4, with the intensity distributions shown in Extended Data Fig. 3i. For example, we observe a broad distribution of emission intensity peaking at 2×10^4 counts for $x=0.20$ compared to a narrower distribution but with a maximum at only 5×10^3 counts for the reference film ($x=0$). Although the intensity distribution is broader for the $x=0.2$ sample, the lowest counts for this sample are as high in number as the highest counts for the reference. We find that the absolute emission intensity increases continuously with increasing content of potassium (Extended Data Fig. 3i), peaking at $x=0.40$ with an order of magnitude enhancement compared with the reference, consistent with the PLQE data in Fig. 1a. We also show the centre-of-mass photoluminescence wavelength maps for the same perovskite films in Extended Data Fig. 3e–h, along with the corresponding histograms in Extended Data Fig. 3j. This further confirms the redshifting of the microscale photoluminescence upon addition of potassium, consistent with the absorption and bulk photoluminescence data. Notably, we find a statistically significant correlation between the photoluminescence wavelength and the intensity for each local site (that is, the most emissive sites are the most redshifted, see Extended Data Fig. 3k). This is consistent with the bulk absorption and photoluminescence data, in which the photoluminescence intensity increases and redshifts with increasing potassium content.

Time-resolved photoluminescence. Time-resolved photoluminescence measurements were acquired with a gated intensified CCD camera system (Andor iStar DH740 CCI-010) connected to a grating spectrometer (Andor SR303i). Excitation was performed with femtosecond laser pulses which were generated in a home-built set-up by second harmonic generation in a β -barium borate crystal from the fundamental output (pulse energy 1.55 eV, pulse length 80 fs) of a Ti:sapphire laser system (Spectra-Physics Solstice). Temporal resolution of the photoluminescence emission was obtained by measuring the photoluminescence from the sample by stepping the intensified CCD gate delay relative to the pump pulse. The gate width was 20 ns.

For the unpassivated control film ($x=0$), there is a considerable initial non-radiative component (Fig. 1b). With increasing passivation, this component is gradually removed with a peak 'lifetime' occurring at $x=0.1$. In this regime, the recombination is still trap-limited for this excitation density (that is, the carrier density does not exceed the trap density to fill all traps). As the passivation increases, the trap density is further reduced, until the density of excited carriers is greater than the trap density and recombination is radiative and bimolecular. In this regime, the recombination follows a power law and the lifetime is 'faster' with decreasing trap density (increasing passivation, that is, $x=0.2$ –0.4). We note that the total area under the non-normalized curves in each case matches the increasing PLQE trend.

Time-resolved microwave conductivity measurements. The TRMC technique monitors the change in microwave power reflected by the loaded microwave cavity upon pulsed laser excitation. The photoconductance (ΔG) of the perovskite films was deduced from the collected laser-induced change in normalized microwave power $\Delta P(t)/P = -K\Delta G(t)$, where K is the sensitivity factor. The product of the yield of generated free charges ϕ and mobility $\Sigma\mu = (\mu_e + \mu_h)$ were obtained by: $\phi\Sigma\mu = \Delta G/(I_0\beta eF_A)$, where I_0 is the number of photons per pulse per unit area, β is a geometry constant of the microwave cell, e is the elementary charge and F_A is the fraction of light absorbed by the sample at the excitation wavelength of 600 nm.

In Extended Data Fig. 5, we show the ΔG as a function of time after pulsed excitation at 600 nm for different fractions of potassium acquired at different fluences. The TRMC signal depicts a rapid increase, originating from the formation of mobile charges, followed by a decay attributed to charge recombination or immobilization of charge carriers by traps. In Fig. 1c, we show that at least up to $x=0.1$, the charge transport remains unperturbed with the addition of potassium. That is, in view of the low electric field strength of the microwaves (100 V cm^{−1}), the charges are not effectively displaced by this field but instead they are perturbed in their diffusional motion. As a result, the distance R crossed by the charges is limited by the charge-carrier diffusion coefficient and half the oscillation period ν (8.5 GHz), which amounts to:

$$\begin{aligned} R &= \sqrt{\mu(k_B T/e) \frac{1}{2} \nu^{-1}} \\ &= \sqrt{42 \times 0.0259 \times 0.5 \times (8.5 \times 10^9)^{-1}} \\ &= 8.0 \times 10^{-6} \text{ cm} \\ &= 80 \text{ nm} \end{aligned}$$

As shown in Extended Data Fig. 1, the grain sizes are approximately 200–400 nm and are independent of the potassium concentration for values up to $x=0.1$. Therefore, for $x < 0.1$, the grains are larger than the probing length (80 nm) of the measurement and the TRMC mobility as probed will not be affected by the grain

size. However, smaller grain sizes are observed for $x > 0.2$, which is most likely the reason for the lower mobilities at higher values of x (ref. 19).

Furthermore, we show the resulting half-lives $\tau_{1/2}$ (time taken to decay to half of the initial value) for the different samples in Extended Data Fig. 5f across a range of excitation fluences. At low fluence, one of the carriers is trapped, leading to a long-lived signal from the untrapped carrier, which recombines following monomolecular kinetics^{13,35}. At higher fluences in which the traps are filled, the recombination is bimolecular. We find that charge-carrier recombination is substantially slower for the $x = 0.1$ composition compared to the $x = 0$ reference, with the low-fluence monomolecular lifetime increasing from 1 μ s ($x = 0$) to 1.5 μ s ($x = 0.1$).

Scanning transmission electron microscopy–energy dispersive X-ray spectroscopy. A FEI Helios Nanolab dual beam focus ion beam/field emission gun-scanning electron microscope (FIB/FEGSEM) was used to prepare a lamella for STEM imaging and analysis. To preserve the perovskite film during specimen preparation, capping layers of spiro-OMeTAD and platinum were deposited. All imaging was carried out in STEM-HAADF (high-angle annular dark field) mode. STEM–EDX data were acquired in an FEI Tecnai Osiris TEM equipped with a high-brightness Schottky X-FEG gun and a Super-X EDX system composed of four silicon drift detectors, each approximately 30 mm² in area and placed symmetrically around the optic axis to achieve a collection solid angle of 0.9 sr. Spectrum images were acquired with a probe current of 0.7 nA, an acceleration voltage of 200 kV, a spatial sampling of 10 nm per pixel and a dwell time of 100 ms per pixel. Data were acquired with a Tecnai Imaging and Analysis system and analysed with Hyperspy.

In Extended Data Fig. 6b–e, we report a cross-section STEM–EDX measurement for a control perovskite thin film (that is, $x = 0$). We treated the EDX dataset with the NMF algorithm, which led to the identification of two main components in the thin film. Factor 1 is representative of the perovskite layer, whereas factor 2 can be associated with a bromide- and lead-rich phase. This second component shows higher intensity at the interfaces with the spiro-OMeTAD protective capping layer and with the silicon substrate. In contrast to the case of the passivated perovskite film ($x = 0.20$, Fig. 3), we do not observe the presence of surface decoration at the grain boundaries.

X-ray diffraction. X-ray diffraction was performed using a Bruker X-ray D8 Advance diffractometer with Cu K $\alpha_{1,2}$ radiation ($\lambda = 1.541$ Å). Spectra were collected with an angular range of $5^\circ < 2\theta < 60^\circ$ and $\Delta\theta = 0.01227^\circ$ over 10 min. Measurements were made on as-prepared films. A Le Bail analysis was carried out on film measurements using the Bruker Topas software. Chebyshev polynomials were used to fit the background and the peak shape modelled with a pseudo-Voigt function.

Grazing incidence wide angle X-ray scattering. GIWAXS measurements were performed on the XMaS facility at the ESRF synchrotron. A fixed-exit, water-cooled, double crystal Si(111) monochromator, placed 25 m from the source, was used to monochromatize the X-ray beam coming from a bending magnet ($E_c = 9.8$ keV). The X-ray energy was tuned to 10 keV (1.2398 Å) and a Rh-coated toroidal mirror was used to focus the monochromatic beam horizontally and vertically. The beam flux was approximately 5×10^{10} photons per second at the sample position. The original beam spot size was 500 (horizontal) \times 400 (vertical) μ m² at the sample position. We used a set of motorized slits (Huber) immediately before the sample to have a better-defined footprint in the vertical direction. The final beam spot size with slits was 300 (horizontal) \times 115 (vertical) μ m². The beam footprint extended 300 μ m horizontally and throughout the perovskite films. The samples were scanned at an out-of-plane incident angle of around 0.3° .

As shown in Extended Data Fig. 7a, d ($x = 0$ and $x = 0.2$, respectively), the GIWAXS diffraction patterns collected at an incident angle of 0.3° show the main perovskite diffraction ring at $q = 1$ Å^{−1}. A PbI₂ peak can be detected at $q = 0.9$ Å^{−1} for the reference film, which probably originates from a small fraction of excess lead iodide in the reference precursor solution. In Extended Data Fig. 7f, g, we plot the line profiles, azimuthally integrated over the entire image for perovskite films with $x = 0$ –0.4. The PbI₂ peak diminishes with higher potassium content and disappears at $x = 0.2$. Furthermore, we observe a new diffraction peak at $q = 0.7$ Å^{−1} that appears when $x \geq 0.10$ and can be assigned to a non-perovskite potassium-rich crystalline phase which grows with increasing potassium. We also observe the growth of new peaks with increasing potassium content in laboratory X-ray diffraction results (highlighted by asterisks in Extended Data Fig. 2d). Currently, we cannot unambiguously assign these peaks—for example to potassium bromide or potassium iodide—which points to this phase (or phases) having more complicated chemical compositions (such as KBr_{*x*}I_{*1−x*}, K₂PbI₄ or K₂PbBr₄²⁴) and crystallinities. However, given the observations in the STEM–EDX experiments (Fig. 3), we expect that at least one of the phases present is rich in potassium and

bromide and is likely to be the halide-sequestering species proposed in our model. Identification of the precise composition of these sequestering species should be the subject of future work in this area.

Soft and hard X-ray photoelectron spectroscopy. Photoelectron spectroscopy measurements were performed at the Diamond Light Source using the beamline I09 (Oxfordshire, UK). Both soft (758 eV) and hard (2,200 and 6,600 eV) X-rays were used to illuminate our sample, to provide information from different probing depths. The soft X-ray energy was selected through a plane grating monochromator, whereas a double-crystal monochromator was used in the hard X-ray section. The values of 2,200 and 6,600 eV correspond to the first- and third-order light when the double-crystal monochromator is set at 2,200 eV using a Si(111) crystal. A EW4000 photoelectron analyser (VG Scienta) was used to record the spectra with an analyser slit open to 0.2 mm. No charge neutralization was used and the binding energy scale was calibrated by setting the Au 4f core level of a gold reference sample to 84.0 eV. The quantification tables and intensity ratios were calculated from the experimental results after correction using the photoionization cross-sections for each element at their specific photon energies, according to database values.

The overview spectra of the perovskite thin films shows all the core-level peaks corresponding to the elements forming the perovskite material, including Cs. The high-resolution spectra of the core-level peaks further confirm the quality of the perovskite films, as there is only a single Pb 5d component without any metallic lead feature³⁶ (Extended Data Fig. 8). In Extended Data Fig. 8, we summarize a quantitative analysis of the potassium content in perovskite films with $x = 0$ –0.2 at different photon energies of 758, 2,200 and 6,600 eV, which corresponds to moving from probing the surface (around 2–4 nm) to probing further into the bulk (around 20 nm). We observed that the concentration of potassium decreases when moving from the surface into the bulk for each of the potassium compositions. Notably, we found that the caesium content increases when moving into the bulk, concomitant with the decrease in potassium, suggesting that the potassium pushes the caesium further into the bulk (Extended Data Fig. 8d). Furthermore, the I/Pb ratio increases at higher potassium content (Extended Data Fig. 8e), particularly at the surface, consistent with the addition of excess I from the potassium iodide source and in agreement with a previous report showing a higher concentration of iodide at the surface³⁶. Finally, the Br/I ratio decreases at higher potassium content, particularly deeper in the film (probed with the higher photon energies), consistent with potassium more selectively drawing out the bromide from the lattice (Extended Data Fig. 8f). These results suggest that potassium is predominantly located at the surfaces, and the addition of potassium iodide also leads to excess halides on the surfaces with a more selective interaction with bromides.

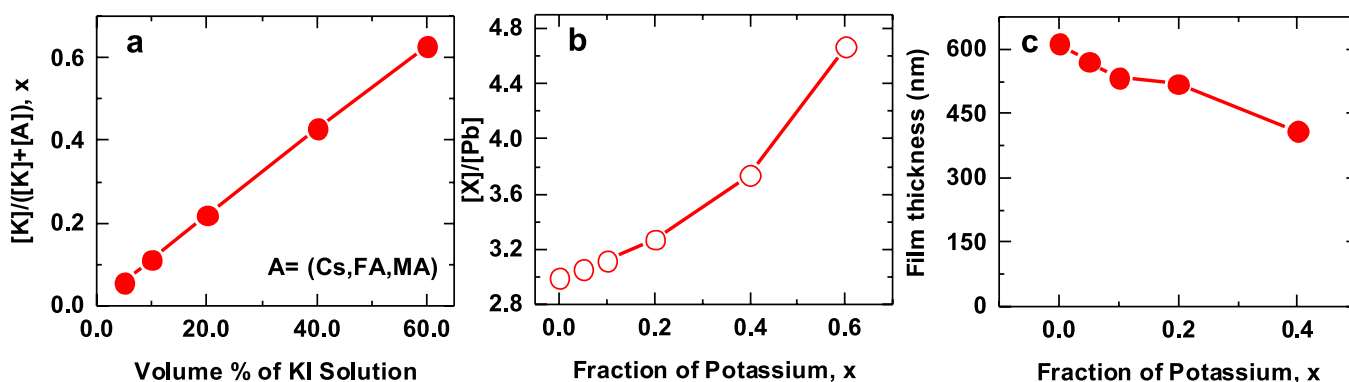
Solar-cell characterization. Current–voltage characteristics were recorded by applying an external potential bias to the cell while recording the generated photocurrent with a digital source meter (Keithley Model 2400). The light source was a 450-W xenon lamp (Oriel) equipped with a Schott-K113 Tempax sunlight filter (Präzisions Glas & Optik GmbH) to match the emission spectrum of the lamp to the AM1.5G standard. Before each measurement, the exact light intensity was determined using a calibrated silicon reference diode equipped with an infrared cut-off filter (KG-3, Schott). External quantum efficiency spectra were recorded as a function of wavelength under a constant white-light bias of approximately 5 mW cm^{−2} supplied by an array of white light-emitting diodes. The excitation beam coming from a 300 W xenon lamp (ILC Technology) was focused through a Gemini-180 double monochromator (Jobin Yvon Ltd) and chopped at approximately 2 Hz. The signal was recorded using a Model SR830 DSP lock-in amplifier (Stanford Research Systems). All measurements were conducted using a non-reflective metal aperture of 0.105 cm² to define the active area of the device and avoid light scattering through the sides.

For stability measurements, the solar cells were transferred to a sealable device holder under nitrogen glove-box conditions. During testing the device holder was continuously purged with dry nitrogen, which had been pre-filtered (SGT Super Clean) to minimize residual oxygen, moisture and hydrocarbon content. A Newport solar simulator with equivalent AM 1.5 G 1-sun output was used to illuminate the entire device substrate; short wavelengths were filtered using a 435 nm long-pass filter (Thor Labs FGL435). Ageing under these conditions resulted in an ambient temperature inside the chamber of approximately 40 °C, which was measured by a thermistor next to the solar-cell device and was reached within 30 min of the experiment commencing. Photocurrent characteristics were recorded close to the maximum power point voltage (as ascertained from an initial *J*–*V* curve) using a Keithley 2636 SMU and a custom-written LabView VI code. Devices were stored in a nitrogen-filled glove box in the dark between shelf-life measurements.

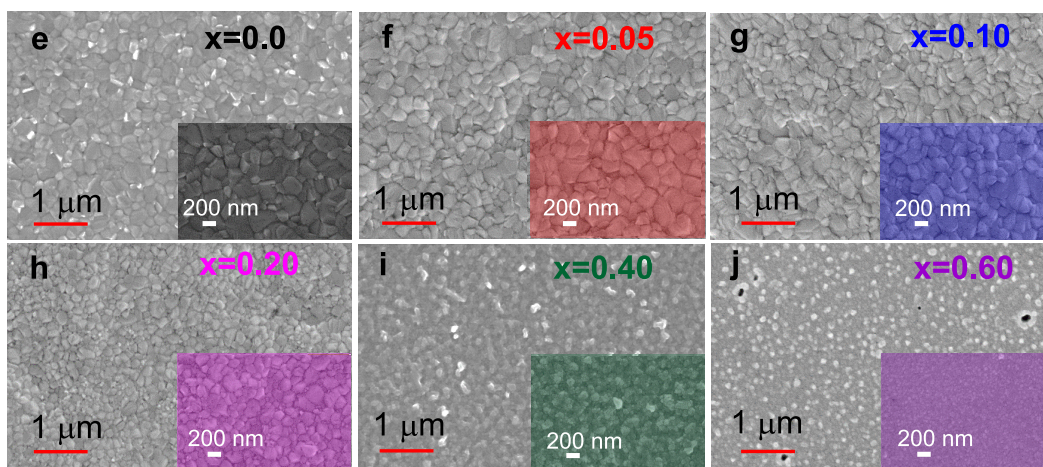
Code availability. The Hyperspy and LabView VI codes used in this work are available from the corresponding author upon reasonable request.

Data availability. The data that support the findings of this study are available from the corresponding author on reasonable request, and at <https://www.repository.cam.ac.uk/handle/1810/271801>.

32. Abdi-Jalebi, M. *et al.* Impact of a mesoporous titania–perovskite interface on the performance of hybrid organic–inorganic perovskite solar cells. *J. Phys. Chem. Lett.* **7**, 3264–3269 (2016).
33. de Mello, J. C., Wittman, H. F. & Friend, R. H. An improved experimental determination of external photoluminescence quantum efficiency. *Adv. Mater.* **9**, 230–232 (1997).
34. Park, B. *et al.* Chemical engineering of methylammonium lead iodide/bromide perovskites: tuning of opto-electronic properties and photovoltaic performance. *J. Mater. Chem. A* **3**, 21760–21771 (2015).
35. Hutter, E. M., Eperon, G. E., Stranks, S. D. & Savenije, T. J. Charge carriers in planar and meso-structured organic–inorganic perovskites: mobilities, lifetimes, and concentrations of trap states. *J. Phys. Chem. Lett.* **6**, 3082–3090 (2015).
36. Philippe, B. *et al.* Chemical distribution of multiple cation (Rb⁺, Cs⁺, MA⁺, and FA⁺) perovskite materials by photoelectron spectroscopy. *Chem. Mater.* **29**, 3589–3596 (2017).



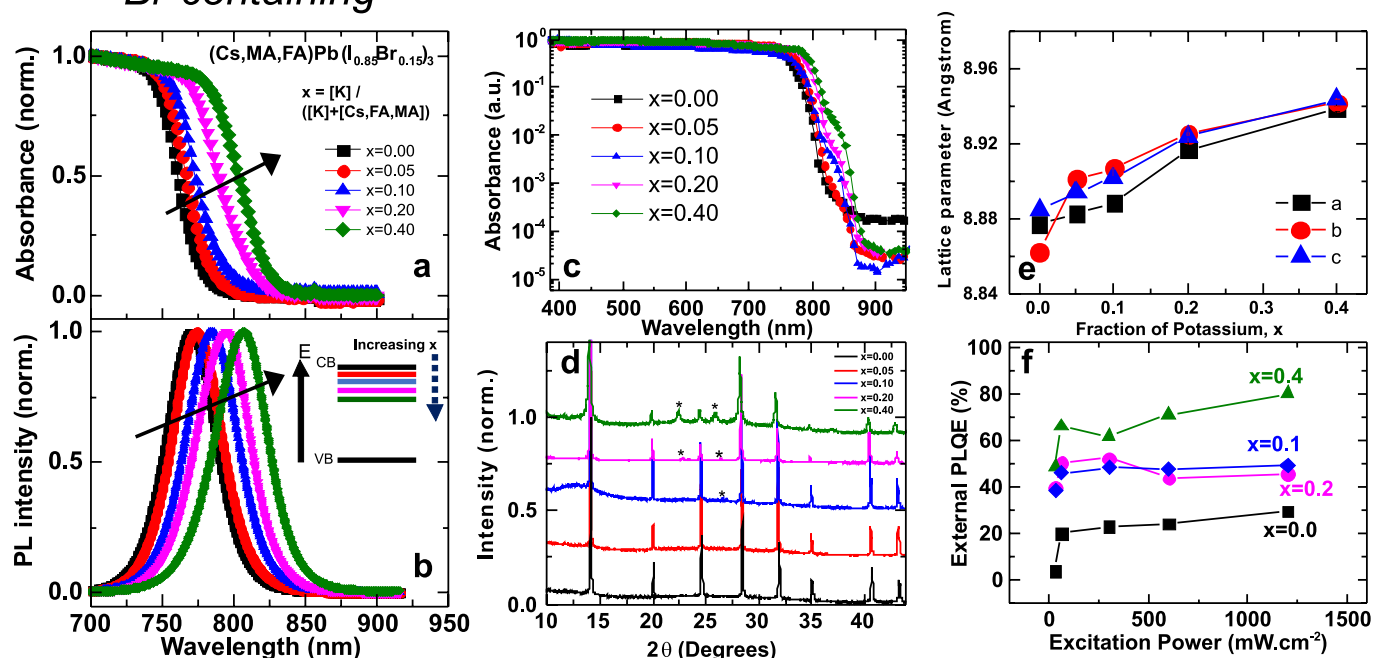
| d Calculated | $x=0.0$ | $x=0.05$ | $x=0.10$ | $x=0.20$ | $x=0.40$ |
|--------------------------------------|---------|----------|----------|----------|----------|
| $[\text{I}]/[\text{Pb}]$ | 2.545 | 2.604 | 2.669 | 2.825 | 3.292 |
| $[\text{Br}]/[\text{X}]$ | 0.149 | 0.146 | 0.143 | 0.137 | 0.120 |
| $[\text{I}]/[\text{X}]$ | 0.851 | 0.854 | 0.857 | 0.863 | 0.880 |
| $[\text{X}]/[\text{Pb}]$ | 2.992 | 3.051 | 3.117 | 3.272 | 3.739 |
| $[\text{K}]/([\text{A}]+[\text{K}])$ | 0 | 0.056 | 0.111 | 0.220 | 0.429 |



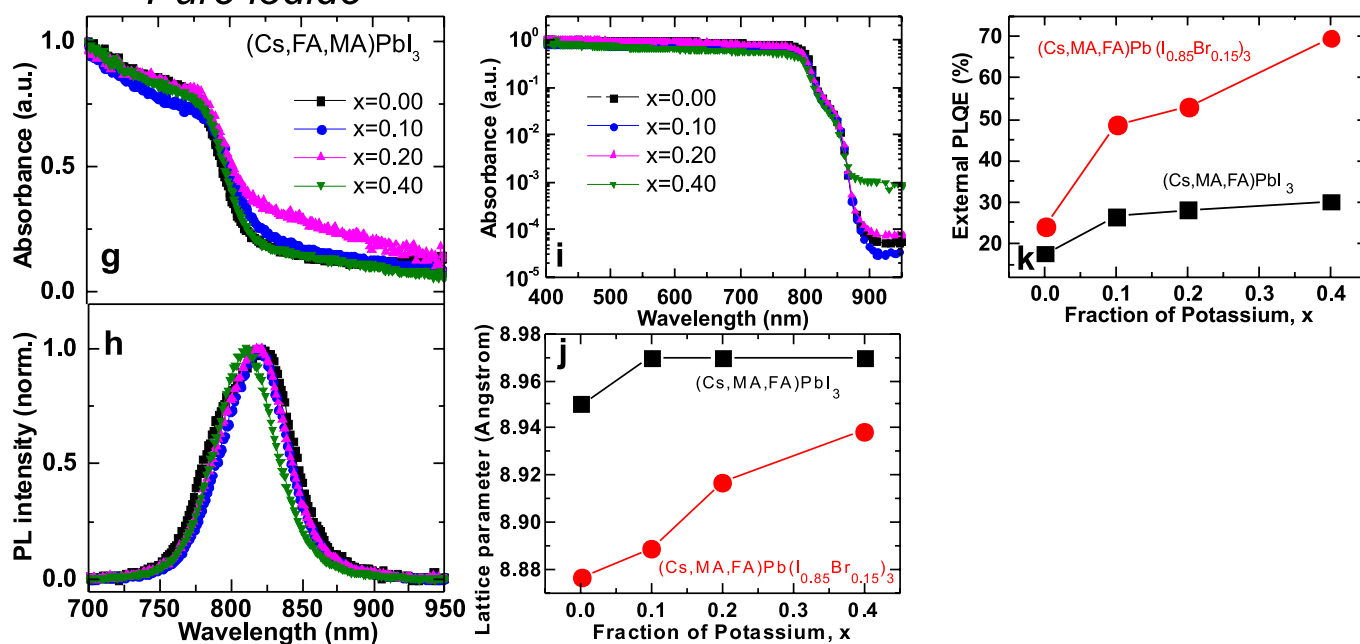
Extended Data Figure 1 | Stoichiometries of the precursor solutions, and thickness measurements and scanning electron microscopy images of the (Cs,FA,MA)Pb(I_{0.85}Br_{0.15})₃ films. **a**, Calculated values of $[K]/([K]+[A])$ as a function of potassium iodide volume ratio added to the (Cs,FA,MA)Pb(I_{0.85}Br_{0.15})₃ perovskite precursor solution. **b**, The halide to lead ratio $([X]/[\text{Pb}])$ for different fractions of potassium (x) in perovskite films based on stoichiometric calculations. **c**, The thickness of

the films deposited on compact-TiO₂/thin-mesoporous TiO₂ electrodes for different fractions of potassium. **d**, The calculated ratios between different elements in the (Cs,FA,MA)Pb(I_{0.85}Br_{0.15})₃ films at different fractions of potassium (x); note that X represents halide (that is, the sum of iodide and bromide). **e–j**, Scanning electron micrographs of (Cs,FA,MA)Pb(I_{0.85}Br_{0.15})₃ perovskite thin films, with potassium fractions (x) ranging from 0.0 to 0.6. The coloured insets show images at higher magnification.

Br-containing

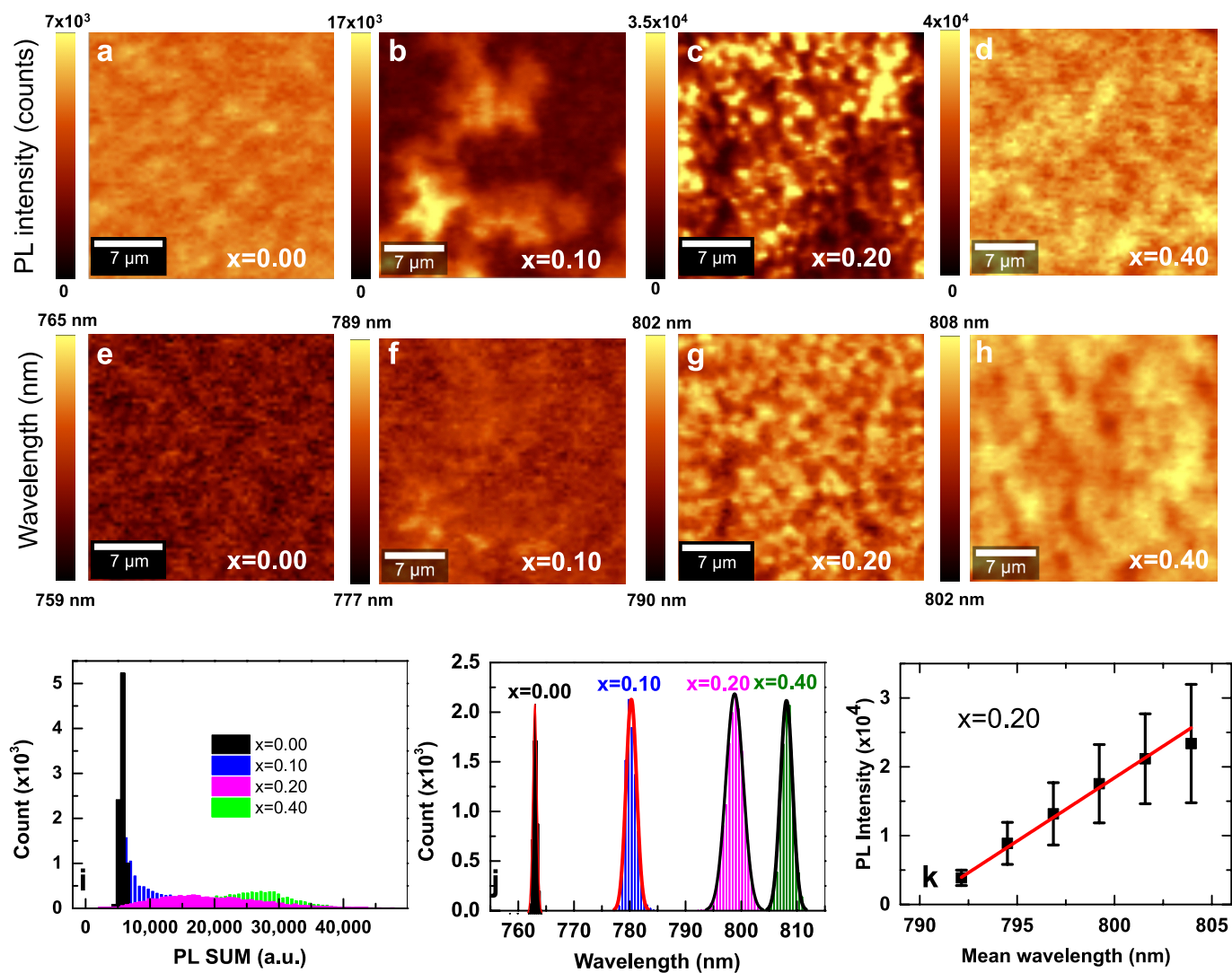


Pure iodide



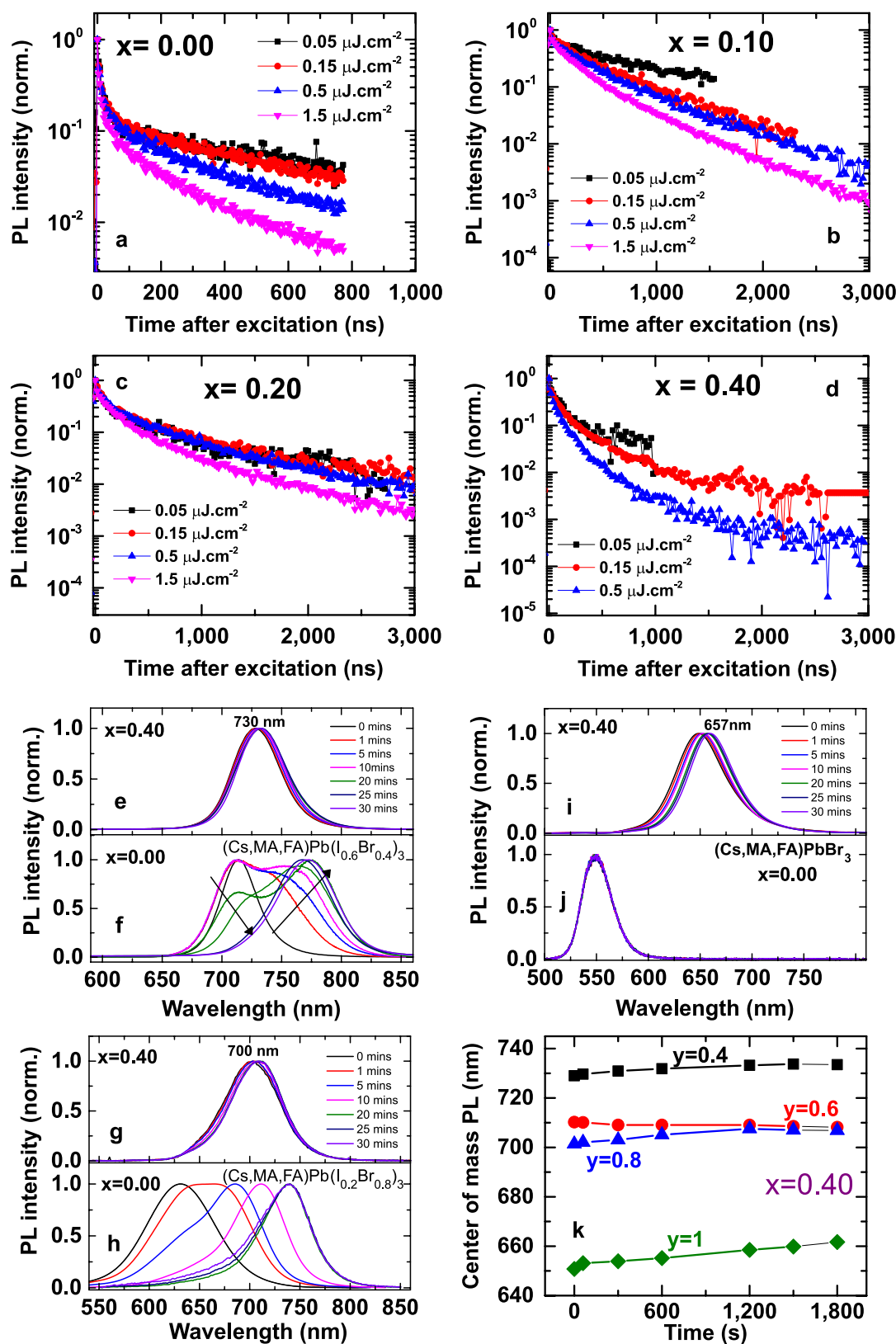
Extended Data Figure 2 | Absorption, photoluminescence and X-ray diffraction characteristics of the $(\text{Cs,FA,MA})\text{Pb}(\text{I}_{0.85}\text{Br}_{0.15})_3$ (bromide-containing) and $(\text{Cs,FA,MA})\text{PbI}_3$ (pure-iodide) films. **a, b**, Normalized UV-vis absorption (**a**) and photoluminescence (PL, 532-nm continuous-wave excitation) (**b**) of the $(\text{Cs,FA,MA})\text{Pb}(\text{I}_{0.85}\text{Br}_{0.15})_3$ perovskite films with different fractions of potassium (x). The inset of **b** is a schematic of the change in energy level of the conduction band with increasing values of x . CB, conduction band; VB, valence band. **c**, The absorption spectra of $(\text{Cs,FA,MA})\text{Pb}(\text{I}_{0.85}\text{Br}_{0.15})_3$ perovskite thin films with different potassium fractions measured by photothermal deflection spectroscopy, showing a decreased sub-gap density of states with potassium passivation. **d**, X-ray diffraction spectra of the $(\text{Cs,FA,MA})\text{Pb}(\text{I}_{0.85}\text{Br}_{0.15})_3$ thin films, with new peaks arising from increasing

potassium labelled with an asterisk. **e**, Calculated lattice parameters using a Le Bail analysis on the X-ray diffraction data. **f**, PLQE as a function of excitation power measured by a 532-nm continuous-wave laser for $(\text{Cs,FA,MA})\text{Pb}(\text{I}_{0.85}\text{Br}_{0.15})_3$ perovskite thin films in an ambient atmosphere. **g–i**, Normalized UV-vis absorption (**g**), photoluminescence (532-nm continuous-wave excitation) (**h**) and photothermal deflection (**i**) spectra of the $(\text{Cs,FA,MA})\text{PbI}_3$ perovskite films with different potassium content. **j**, Calculated lattice parameters, determined using a Le Bail analysis on X-ray diffraction data, for the $(\text{Cs,FA,MA})\text{PbI}_3$ perovskite thin films (black squares) compared to $(\text{Cs,FA,MA})\text{Pb}(\text{I}_{0.85}\text{Br}_{0.15})_3$ (red circles). **k**, PLQE of different passivated perovskite thin films with and without bromide measured under illumination with a 532-nm laser at an excitation intensity equivalent to approximately 1 sun (60 mW cm^{-2}) in ambient air.



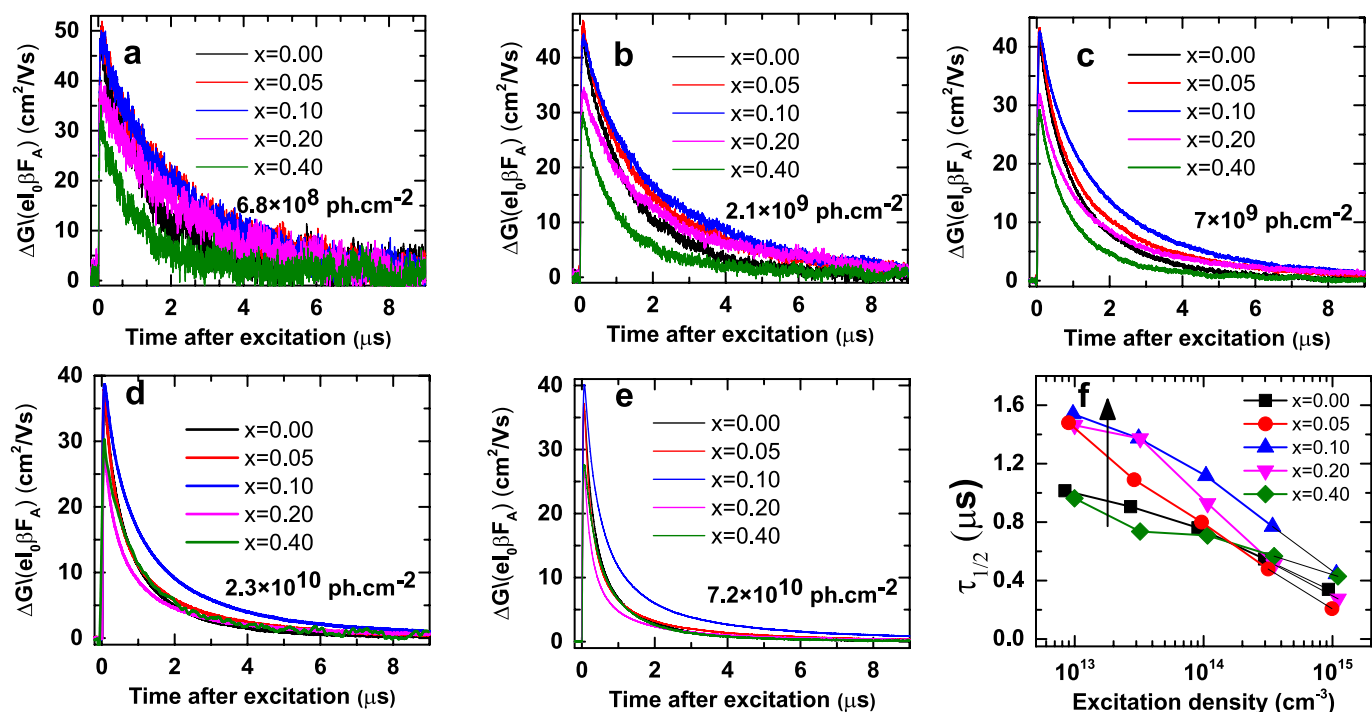
Extended Data Figure 3 | Confocal photoluminescence maps of passivated $(\text{Cs,FA,MA})\text{Pb}(\text{I}_{0.85}\text{Br}_{0.15})_3$ films. **a–d,** Confocal photoluminescence intensity maps with 405-nm excitation measured in ambient atmosphere for $(\text{Cs,FA,MA})\text{Pb}(\text{I}_{0.85}\text{Br}_{0.15})_3$ perovskite thin films with $x = 0.0$ (a), $x = 0.10$ (b), $x = 0.20$ (c) and $x = 0.40$ (d). **e–h,** Centre-of-mass photoluminescence wavelength of the films

for $x = 0.0$ (e), $x = 0.10$ (f), $x = 0.20$ (g) and $x = 0.40$ (h). **i, j,** Histograms of the absolute photoluminescence intensities (i) and photoluminescence wavelength (j) extracted from the respective maps for $x = 0–0.40$. **k,** Correlation between the local photoluminescence intensity and mean wavelength for $x = 0.20$.



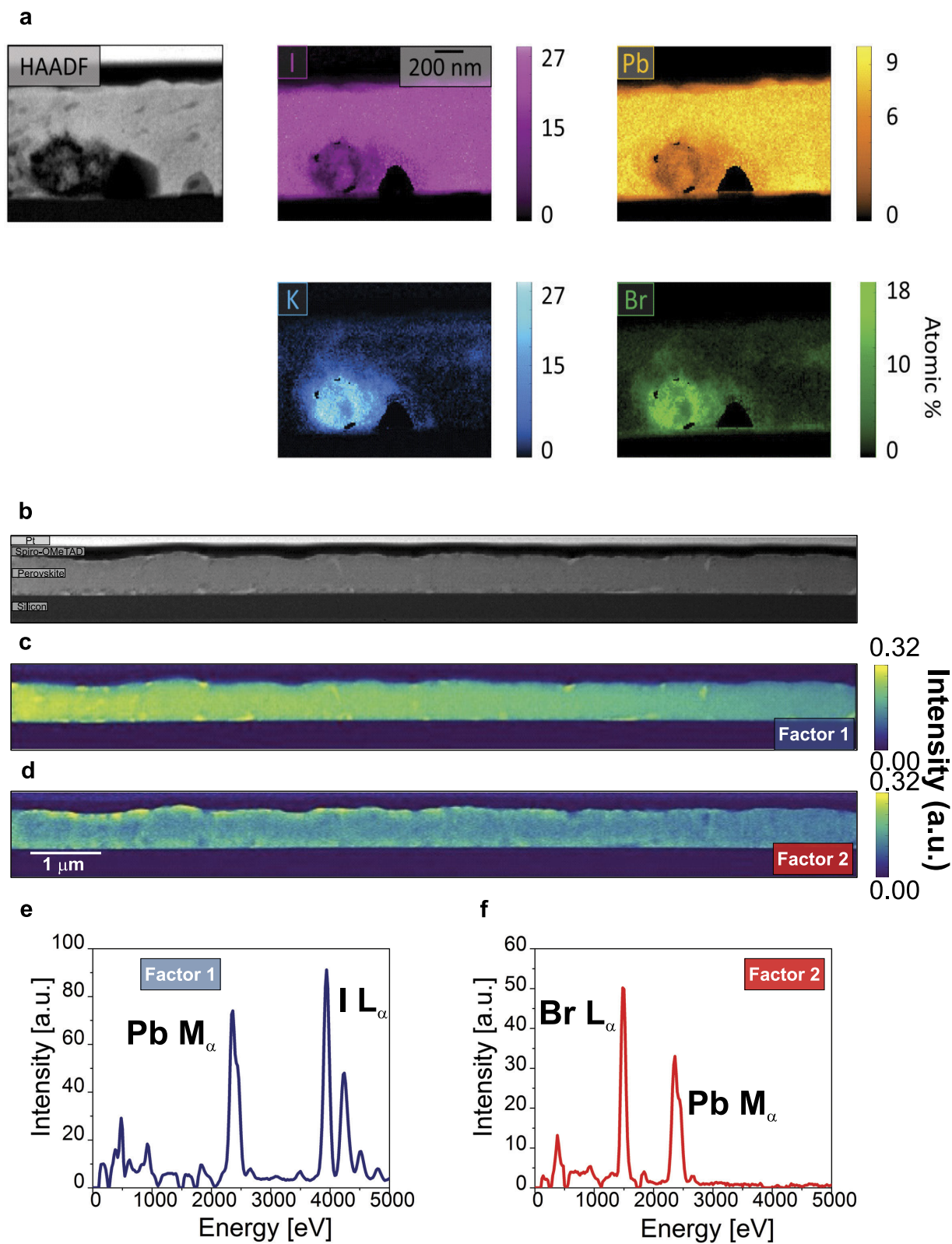
Extended Data Figure 4 | Time-resolved photoluminescence measurements and photostability of passivated (Cs,FA,MA)Pb(I_{1-y}Br_y)₃ films. **a–d**, Intensity dependent time-resolved photoluminescence decays of the (Cs,MA,FA)Pb(I_{0.85}Br_{0.15})₃ perovskite films with different fractions of potassium. The pulse fluences of the 407-nm excitation are quoted on the graphs. **e–j**, Photoluminescence from (Cs,FA,MA)Pb(I_{1-y}Br_y)₃ films with (e, f) $y = 0.4$ passivated ($x = 0.4$, e) and unpassivated ($x = 0$,

f), (g, h) $y = 0.8$ passivated ($x = 0.4$, g) and unpassivated ($x = 0$, h), (i, j) $y = 1$ passivated ($x = 0.4$, i) and unpassivated ($x = 0$, j). The samples were illuminated and the photoluminescence acquired continuously with a 532-nm laser at an excitation intensity equivalent to approximately 1 sun (60 mW cm^{-2}) in an ambient atmosphere. **k**, Centre of mass for the photoluminescence wavelength of the passivated perovskite films ($x = 0.4$).



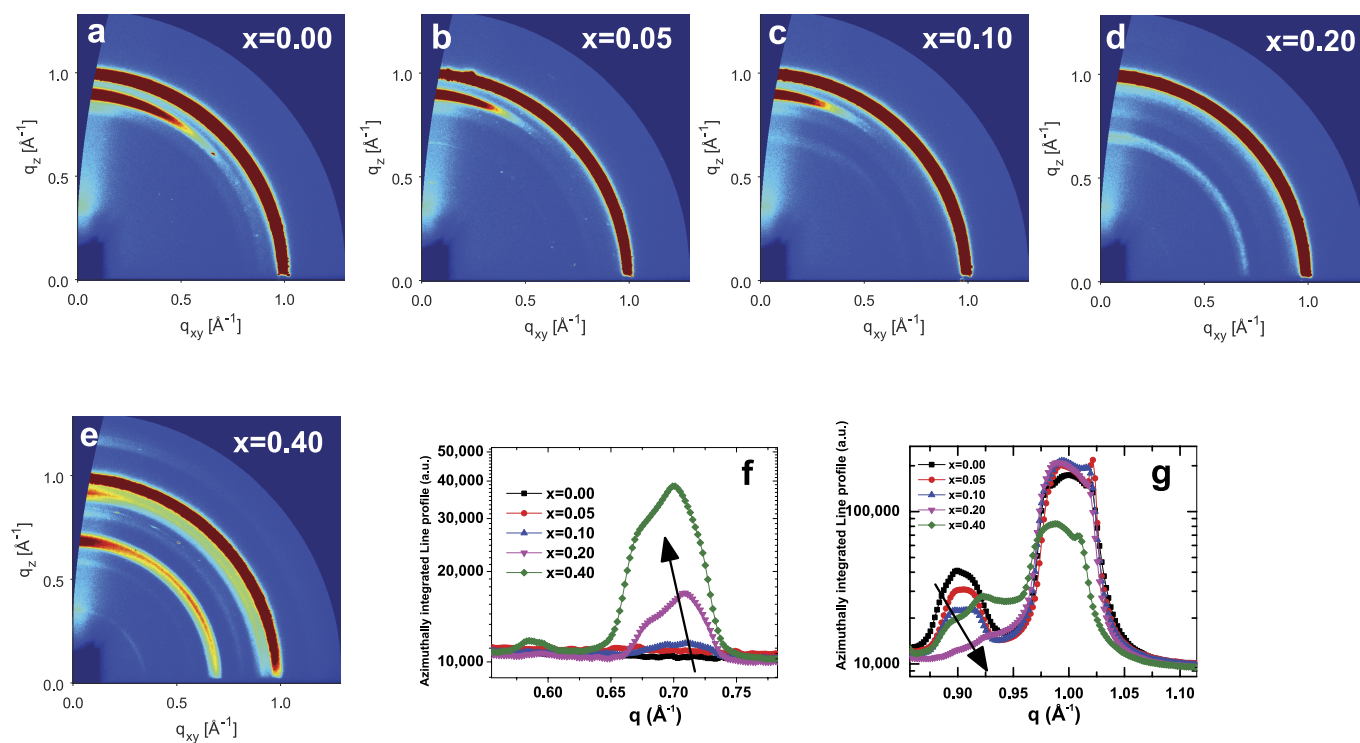
Extended Data Figure 5 | Time-resolved microwave conductivity measurements on passivated (Cs,FA,MA)Pb(I_{0.85}Br_{0.15})₃ films.
a–e, Time-resolved microwave conductivity measurements for (Cs,FA,MA)Pb(I_{0.85}Br_{0.15})₃ perovskite films with different fractions of potassium ($x = 0-0.4$) showing the change in photoconductance after

pulsed excitation at 600 nm, with excitation densities (photons per cm²) as quoted on the graphs. f, Half-life values extracted from the decays, with the excitation density quoted after accounting for the absorbed fraction and film thickness of each sample.



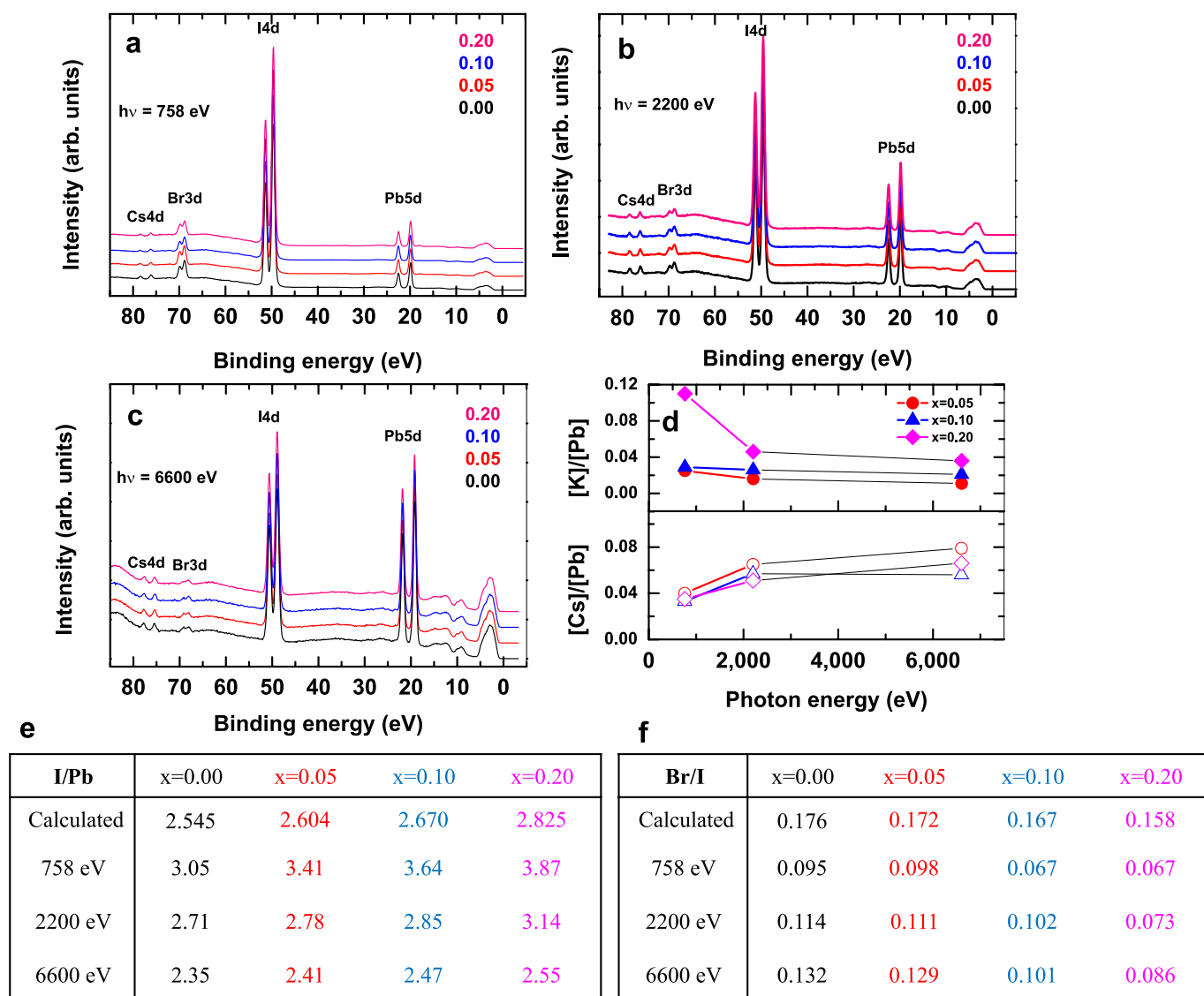
Extended Data Figure 6 | STEM-EDX chemical maps of the passivated and unpassivated (Cs,FA,MA)Pb(I_{0.85}Br_{0.15})₃ samples. a, The HAADF image and corresponding STEM-EDX quantitative maps for iodine, lead, potassium and bromide in a (Cs,FA,MA)Pb(I_{0.85}Br_{0.15})₃ perovskite specimen with a potassium fraction (x) of 0.20. **b,** STEM-HAADF cross-

sectional image of an unpassivated (Cs,FA,MA)Pb(I_{0.85}Br_{0.15})₃ perovskite thin film ($x=0$). **c, d,** NMF decomposition results in factor 1 associated with the perovskite layer (**c**) and in factor 2 indicating the presence of a lead and bromide-rich phase (**d**). **e, f,** STEM-EDX profiles of factor 1 (**e**) and factor 2 (**f**).



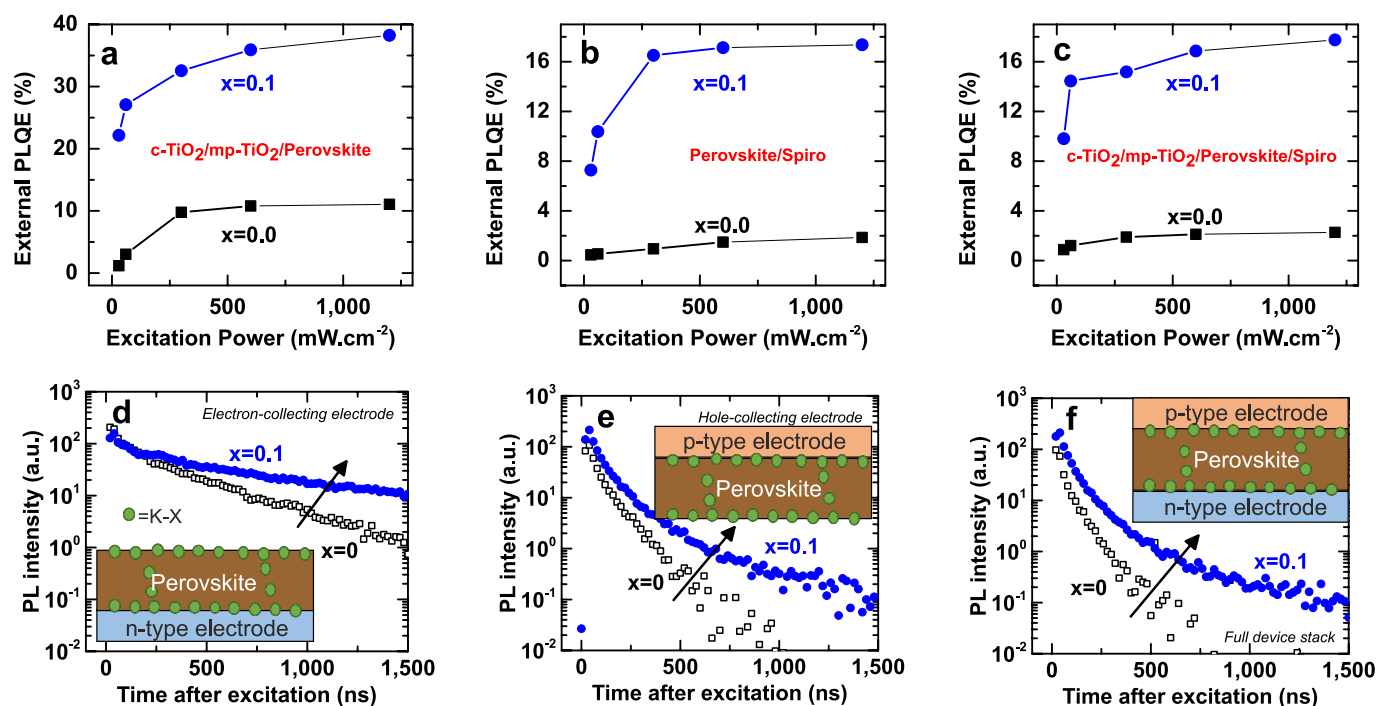
Extended Data Figure 7 | GIWAXS measurements of passivated $(\text{Cs,FA,MA})\text{Pb}(\text{I}_{0.85}\text{Br}_{0.15})_3$ films. a–e, The diffraction patterns of thin $(\text{Cs,FA,MA})\text{Pb}(\text{I}_{0.85}\text{Br}_{0.15})_3$ films collected at low angle using GIWAXS for $x = 0.00$ (a), $x = 0.05$ (b), $x = 0.10$ (c), $x = 0.20$ (d) and $x = 0.40$ (e).

f, g, The high-resolution line profiles azimuthally integrated over the entire GIWAX profile for different fractions of potassium at $0.5 \leq q \leq 0.8$ (f) and $0.8 \leq q \leq 1.1$ (g).



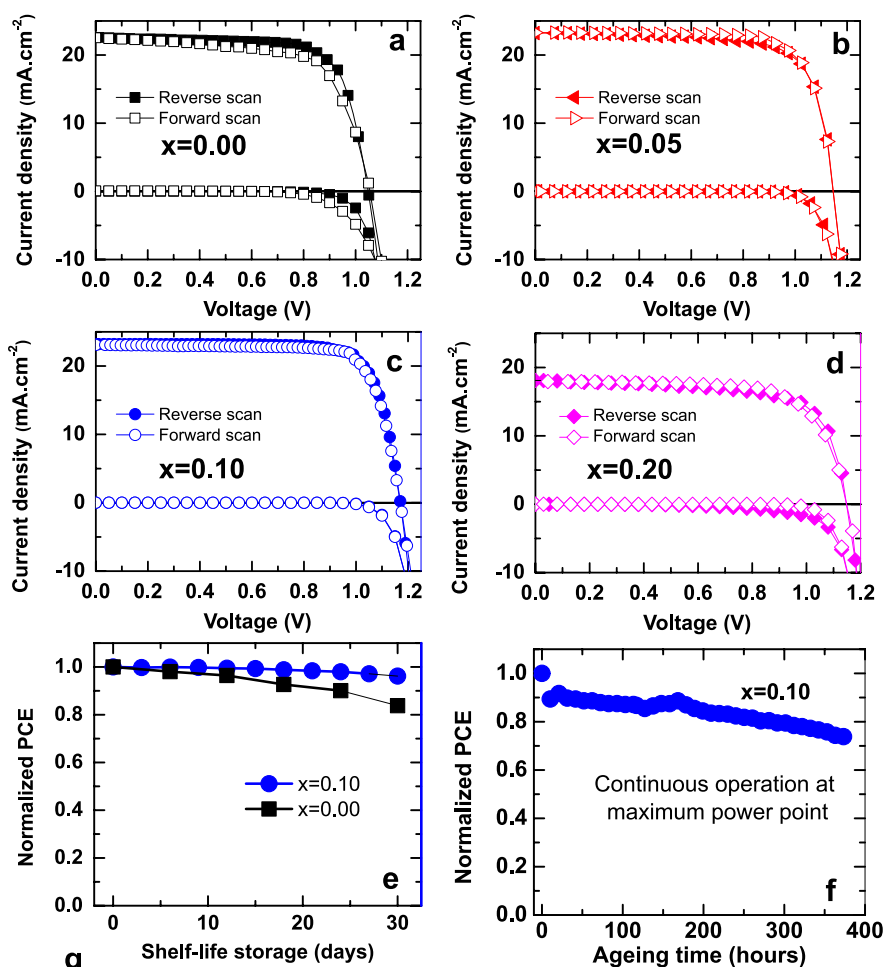
Extended Data Figure 8 | HAXPES spectra for passivated (Cs,FA,MA)Pb(I_{0.85}Br_{0.15})₃ perovskite thin films. a–c, HAXPES spectra for (Cs,FA,MA)Pb(I_{0.85}Br_{0.15})₃ perovskite thin films ($0 \leq x \leq 0.20$) over a binding-energy range of 0 to 85 eV, recorded with a photon energy of 758 eV (a), 2,200 eV (b) and 6,600 eV (c). d, Intensity ratio between core levels ([Cs]/[Pb] and [K]/[Pb]) calculated from the experimental results as

a function of photon energy (measurements at 758, 2,200 and 6,600 eV). e, f, Intensity ratios between different core levels of the perovskite thin films with different fractions of potassium ($0.0 \leq x \leq 0.20$), calculated from experimental results: I/Pb (e) and Br/I (f). We used Pb 5d, K 2p, Cs 4d, I 4d and Br 3d core levels for all energies, with the exception of 6,600 eV, for which K 1s was used.



Extended Data Figure 9 | Excitation-dependent photoluminescence quantum efficiency and time-resolved photoluminescence measurements of (Cs,FA,MA)Pb(I_{0.85}Br_{0.15})₃ device stacks. a–c, PLQE of reference ($x=0$) and passivated ($x=0.1$) (Cs,FA,MA)Pb(I_{0.85}Br_{0.15})₃ perovskite thin films with n-type contact (a), p-type contact (b) and both contacts (c), each measured under illumination with a 532-nm laser at different excitation intensities. d–f, Time-resolved photoluminescence

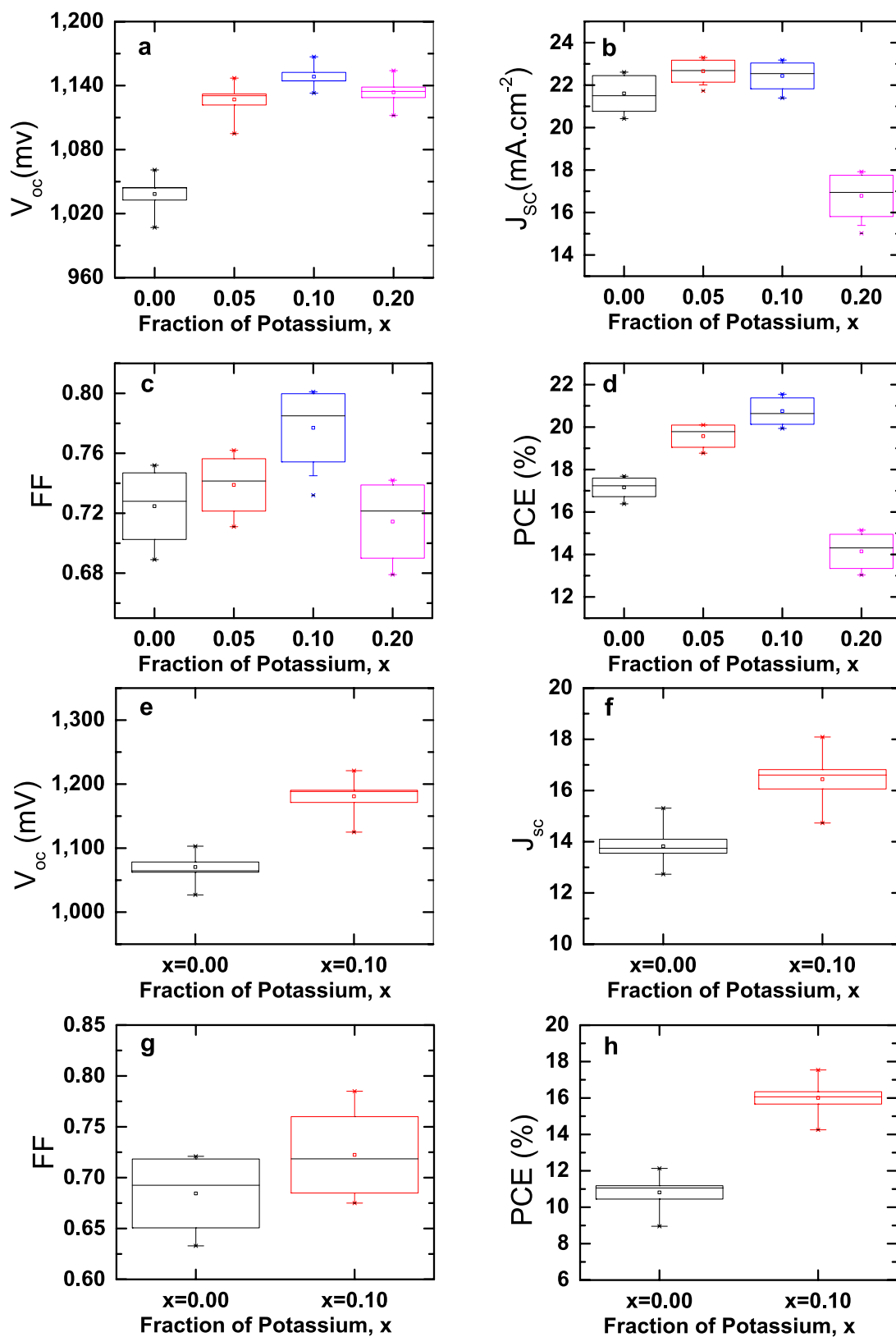
decays of encapsulated (Cs,FA,MA)Pb(I_{0.85}Br_{0.15})₃ films ($x=0$ and $x=0.1$) with excitation at 400 nm and pulse fluence of $0.17 \mu\text{J cm}^{-2}$ ($5 \times 10^{15} \text{ cm}^{-3}$, equivalent to around 3 sun) when the perovskite is interfaced with an n-type electron-collecting electrode (compact-TiO₂/thin-mesoporous TiO₂) (d), a p-type hole-collecting electrode (spiro-OMeTAD) (e), and both electrodes in a full device stack (f).



| x | | J_{sc} [mA cm^{-2}] | V_{oc} [V] | FF | PCE [%] |
|----------------------------------------------------------------|---------|----------------------------------|--------------|------|---------|
| (Cs,FA,MA)Pb($\text{I}_{0.85}\text{Br}_{0.15}$) ₃ | | | | | |
| 0.00 | Reverse | 22.6 | 1.05 | 0.73 | 17.3 |
| | Forward | 22.5 | 1.05 | 0.69 | 16.3 |
| 0.05 | Reverse | 23.3 | 1.14 | 0.75 | 20.1 |
| | Forward | 23.2 | 1.14 | 0.74 | 19.7 |
| 0.10 | Reverse | 23.2 | 1.17 | 0.79 | 21.5 |
| | Forward | 23.1 | 1.17 | 0.79 | 21.3 |
| 0.20 | Reverse | 18.1 | 1.15 | 0.70 | 14.6 |
| | Forward | 17.9 | 1.15 | 0.70 | 14.4 |
| (Cs,FA,MA)Pb($\text{I}_{0.4}\text{Br}_{0.6}$) ₃ | | | | | |
| 0.00 | Reverse | 15.3 | 1.12 | 0.72 | 12.3 |
| | Forward | 15.1 | 1.13 | 0.55 | 9.4 |
| 0.10 | Reverse | 17.9 | 1.23 | 0.79 | 17.5 |
| | Forward | 18.1 | 1.22 | 0.79 | 17.8 |

Extended Data Figure 10 | Current-voltage curves of passivated (Cs,FA,MA)Pb($\text{I}_{0.85}\text{Br}_{0.15}$)₃ devices and tabulated results for the best-performing (Cs,FA,MA)Pb($\text{I}_{0.85}\text{Br}_{0.15}$)₃ and (Cs,FA,MA)Pb($\text{I}_{0.4}\text{Br}_{0.6}$)₃ devices. a–d, Forward (open symbols) and reverse (closed symbols) $J-V$ curves of the best-performing solar cells with (Cs,FA,MA)Pb($\text{I}_{0.85}\text{Br}_{0.15}$)₃ absorbers with $x = 0.00$ (a), $x = 0.05$ (b), $x = 0.10$ (c) and $x = 0.20$ (d) measured under full simulated solar illumination conditions (AM1.5, 100 mW cm^{-2}) with a scan rate of 15 mV s^{-1} . The corresponding dark $J-V$ curves are also shown. e, f, Preliminary stability tests of (Cs,FA,MA)

Pb($\text{I}_{0.85}\text{Br}_{0.15}$)₃ perovskite devices. Shelf-life of devices for $x = 0.00$ and $x = 0.10$ stored in a nitrogen glove box over a month and tested regularly under full AM1.5 simulated sunlight (e), and stability of a device with $x = 0.10$ aged at 0.8 V under continuous ultraviolet-filtered simulated sunlight in a nitrogen atmosphere for over 350 hours (f). g, Device parameters for the passivated (Cs,FA,MA)Pb($\text{I}_{0.85}\text{Br}_{0.15}$)₃ (upper) and (Cs,FA,MA)Pb($\text{I}_{0.4}\text{Br}_{0.6}$)₃ (lower) perovskite solar cells measured under full simulated solar illumination conditions (AM1.5, 100 mW cm^{-2}).



Extended Data Figure 11 | Device statistics. a–h, Box and whisker plots to summarize the statistics of photovoltaic parameters of ten devices with passivated (Cs,FA,MA)Pb(I_{0.85}Br_{0.15})₃ solar cells (a–d) and eight devices of passivated (Cs,FA,MA)Pb(I_{0.4}Br_{0.6})₃ solar cells (e–h), each measured under full simulated solar illumination conditions (AM1.5, 100 mW cm⁻²)

and scanned at a rate of 15 mV s⁻¹. The boxes represent the interquartile range, with the median represented by the line dividing the boxes, and the whiskers are determined by the 5th and 95th percentiles. The mean is given by the open square symbols, and the cross symbols represent the maximum and minimum values.

A lithium–oxygen battery with a long cycle life in an air–like atmosphere

Mohammad Asadi^{1,2}, Baharak Sayahpour¹, Pedram Abbasi¹, Anh T. Ngo³, Klas Karis⁴, Jacob R. Jokisaari⁴, Cong Liu⁵, Badri Narayanan³, Marc Gerard¹, Poya Yasaei¹, Xuan Hu⁴, Arijita Mukherjee⁴, Kah Chun Lau⁶, Rajeev S. Assary³, Fatemeh Khalili-Araghi⁴, Robert F. Klie⁴, Larry A. Curtiss³ & Amin Salehi-Khojin¹

Lithium–air batteries are considered to be a potential alternative to lithium-ion batteries for transportation applications, owing to their high theoretical specific energy¹. So far, however, such systems have been largely restricted to pure oxygen environments (lithium–oxygen batteries) and have a limited cycle life owing to side reactions involving the cathode, anode and electrolyte^{2–5}. In the presence of nitrogen, carbon dioxide and water vapour, these side reactions can become even more complex^{6–11}. Moreover, because of the need to store oxygen, the volumetric energy densities of lithium–oxygen systems may be too small for practical applications¹². Here we report a system comprising a lithium carbonate-based protected anode, a molybdenum disulfide cathode² and an ionic liquid/dimethyl sulfoxide electrolyte that operates as a lithium–air battery in a simulated air atmosphere with a long cycle life of up to 700 cycles. We perform computational studies to provide insight into the operation of the system in this environment. This demonstration of a lithium–oxygen battery with a long cycle life in an air-like atmosphere is an important step towards the development of this field beyond lithium-ion technology, with a possibility to obtain much higher specific energy densities than for conventional lithium-ion batteries.

We used two strategies to limit side reactions in a lithium–oxygen battery in a simulated air atmosphere with representative amounts of O₂, N₂, CO₂ and H₂O. First, we developed a Li₂CO₃/C coating for the lithium anode that allows only lithium cations to pass through, thus protecting the anode from the components of the simulated air. Li₂CO₃ was not expected to react with water to produce the bicarbonate, because the reaction is not thermodynamically favourable under ambient conditions (Supplementary Information section 1). Second, we constructed a cathode based on previously reported molybdenum disulfide nanoflakes², and a mixture of the ionic liquid 1-ethyl-3-methylimidazolium tetrafluoroborate (EMIM-BF₄) and dimethyl sulfoxide (DMSO) was used as the electrolyte. The components of this system operate in conjunction to prevent the formation of side products in the presence of CO₂ and H₂O.

The Li₂CO₃/C anode-protection coating was directly synthesized on a Li anode in a custom-made electrochemical lithium–carbon dioxide cell filled with pure CO₂; ten continuous discharge–charge cycles were operated to form the protective coating (Supplementary Information section 1). Scanning electron microscopy (SEM) images of the coated anode (Fig. 1a and Supplementary Fig. 2a–c) reveal a dense network of rod-shape structures on the surface. Raman spectra show distinct peaks at 717, 743, 1088 and 1,456 cm^{−1} for this coating (Fig. 1b), which are characteristic peaks of Li₂CO₃ (ref. 13) (Supplementary Information section 3). Moreover, X-ray photoelectron spectroscopy (XPS; Fig. 1c, d and Supplementary Fig. 4a–c) for Li 1s, C 1s and O 1s reveals peaks at 55.2, 289.3 and 531.5 eV, respectively, further confirming the presence

of Li₂CO₃ (ref. 14). A peak at 284.8 eV corresponds to bonds between carbon atoms, and is attributed to adventitious carbon compounds as well as solid carbon from the reaction of Li and CO₂ (Supplementary Information section 1). The synthesized Li₂CO₃/C coating was further characterized by electron energy loss spectroscopy (EELS). The K-edge peaks of lithium, oxygen and carbon, obtained from EELS of platelet-like particles extracted from the anode, are presented in Fig. 1e. The sharp peaks corresponding to these elements are similar to those observed for Li₂CO₃ in the solid–electrolyte interphases of lithium-ion batteries¹⁵ (Supplementary Information section 5).

The lithium retention of the protected anode was investigated by running a cell in an air environment for 51 cycles, followed by an exhaustive stripping test, with the results shown in Fig. 1f. This indicated that the protected anode has an average lithium retention of 99.97% per cycle (Supplementary Information section 6). Moreover, electrochemical impedance spectroscopy was used to study the charge-transfer resistance (*R*_{ct}) of the protective layer (Supplementary Information section 7). The results shown in Fig. 1g indicate that the *R*_{ct} of the protected anode (formed from ten deposition cycles) is around 550 kΩ, which is about 20 times greater than that of an unprotected anode (30 kΩ), confirming the existence of an electrically insulating protective coating on the surface of the anode.

A custom-made cell with a MoS₂ cathode, a protected lithium anode and an EMIM-BF₄/DMSO (25%/75%) electrolyte was used in the lithium–air experiments (Supplementary Information sections 8–10). This electrolyte composition provides the maximum oxygen reduction and evolution in a three-electrode electrochemical cell (Supplementary Information section 10). A custom-made simulated air stream of around 79% N₂, around 21% O₂, 500 p.p.m. CO₂, and a relative humidity of 45% at 25 °C was used for the battery experiments (Supplementary Information section 11). Figure 2a shows the long-term discharging and charging profiles up to a capacity of 500 mAh g^{−1} with a constant current density of 500 mA g^{−1}. The charge at the first cycle began at 2.92 V, which is very close to the reversible thermodynamic potential of Li₂O₂ formation (2.96 V versus Li/Li⁺)¹⁶ and reached a potential of 3.75 V at a capacity of 500 mA g^{−1}. The potential gap for the first cycle of the lithium–air system is 0.88 V, increasing to 1.3 V after 50 cycles, followed by a gradual increase to 1.62 V after 550 cycles. The increase in the potential gap during cycling may be due to slow degradation of the protective anode coating and/or the MoS₂ cathode. We did not observe any failure of the battery during testing for up to 700 cycles (Supplementary Information section 12). Figure 2b shows the dependence of the number of discharge–charge cycles achieved in air on the number of deposition cycles used to form the anode-protection layer. The results indicate a substantial increase in the number of lithium–air cycles achieved when the anode is protected compared with when it is not; with no coating, the lithium–air cell

¹Department of Mechanical and Industrial Engineering, University of Illinois at Chicago, Chicago, Illinois 60607, USA. ²Department of Chemical and Biological Engineering, Illinois Institute of Technology, Chicago, Illinois 60616, USA. ³Materials Science Division, Argonne National Laboratory, Argonne, Illinois 60439, USA. ⁴Department of Physics, University of Illinois at Chicago, Chicago, Illinois 60607, USA. ⁵Chemical Sciences Division, Argonne National Laboratory, Argonne, Illinois 60439, USA. ⁶Department of Physics and Astronomy, California State University, Northridge, California 91330, USA.

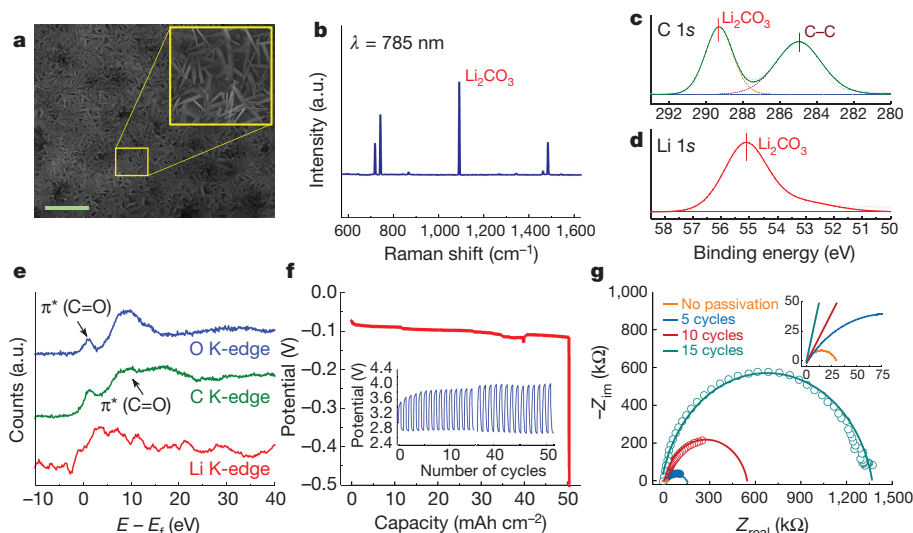


Figure 1 | Characterization of the protected anode. **a**, SEM image of the protected anode surface (scale bar, 1 μm). **b**, Raman spectrum of the protected anode, excited by a 785-nm laser. **c**, **d**, XPS spectra of the protected anode surface in the Li 1s (**c**) and C 1s (**d**) regions. **e**, EELS spectra showing the lithium, carbon and oxygen K-edges of the anode surface, consistent with the presence of Li_2CO_3 . **f**, Exhaustive stripping

of protected lithium using a current density of 0.5 mA cm^{-2} . The inset shows a cycling test of the protected anode in air under a current density of 0.05 mA cm^{-2} . **g**, Electrochemical impedance spectroscopy of the passivated anode after 5, 10 and 15 cycles of exposure to a CO_2 environment, compared to an anode without passivation.

fails after 11 cycles, whereas up to 700 cycles can be achieved with an anode-protection layer. The thickness of the anode-protection layer helps to understand these results. Although the electronic conductivity of thinner coatings can result in electrolyte decomposition, the low ionic conductivity of thicker coatings can result in larger charge potentials and lead to deleterious side reactions. In this case, ten deposition cycles resulted in the optimum thickness to balance these effects.

Raman spectroscopy and XPS were performed on the surface of the cathode to study the discharge products and the cell chemistry after cycling. In terms of lithium species, the Raman spectra (Fig. 2c) show the presence of only a Li_2O_2 peak at 788 cm^{-1} ; we did not detect any peaks related to LiOH , Li_2CO_3 or LiO_2 , and the Li_2O_2 peaks were not present in the charged samples (Supplementary Information section 13).

To study the possibility of any chemical reactions of the electrolyte with Li_2O_2 , we aged a 50-cycle discharged cathode sample in electrolyte for 200 hours; the Raman spectra after ageing did not show any evidence of side products. To further confirm the presence of Li_2O_2 , we performed XPS analysis on the cathode surface. Figure 2d–f shows the Li 1s, C 1s and O 1s spectra of the discharged cathode after 250 cycles. The Li 1s and O 1s peaks of Li_2O_2 , at 55.07 and 531.12 eV, respectively, are in agreement with previously reported *in situ* ambient-pressure XPS studies on Li_2O_2 formed in a lithium–oxygen cell^{17,18}. The C 1s spectra show the reference C–C bond of carbon at 284.8 eV and other carbon peaks at 286.6 and 288.5 eV, which probably arise from the gas diffusion layer. Figure 2d–f also confirms the absence of Li_2CO_3 and LiOH during discharge. Similar XPS results were obtained after one cycle and

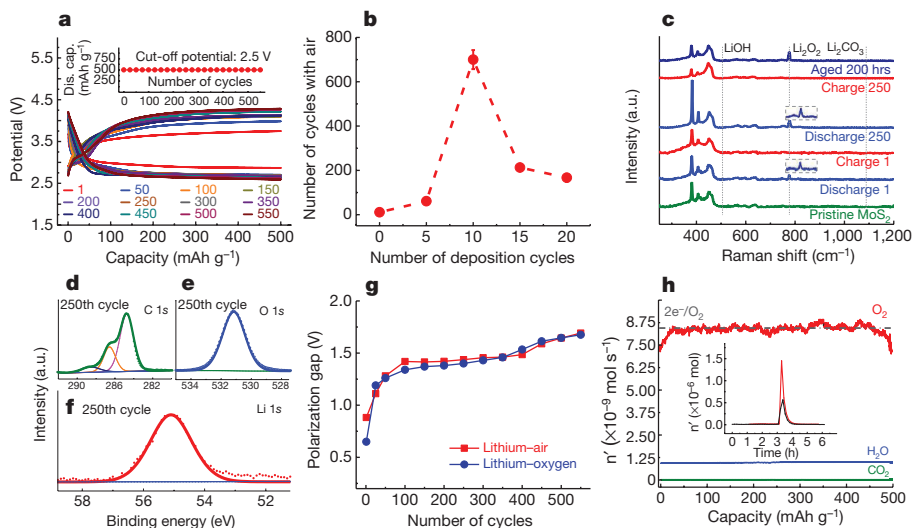


Figure 2 | Performance of the cathode in a lithium–air battery system. **a**, The discharge–charge voltage profile over 550 cycles. The inset shows the capacity versus the number of cycles. **b**, The dependence of the number of battery cycles achieved in air on the number of deposition cycles used to form the anode-protection layer. The error bar shows the standard deviation of three measurements. **c**, Raman spectra of the cathode after the first and after the 250th discharge–charge cycle, compared to pristine MoS_2 and a 200-hour-aged sample. **d–f**, XPS spectra of the cathode surface

for C 1s (**d**), O 1s (**e**) and Li 1s (**f**) after the 250th discharge cycle. **g**, The polarization gap between the lithium–air battery and the Li–O_2 battery, measured under the same operating conditions, as a function of cycle number. **h**, DEMS profiles of the cell during the first charging process, after the cell was discharged to 500 mAh g^{-1} . The inset shows the number of moles of oxygen detected by DEMS, before (red line) and after (black line) discharge in the first cycle.

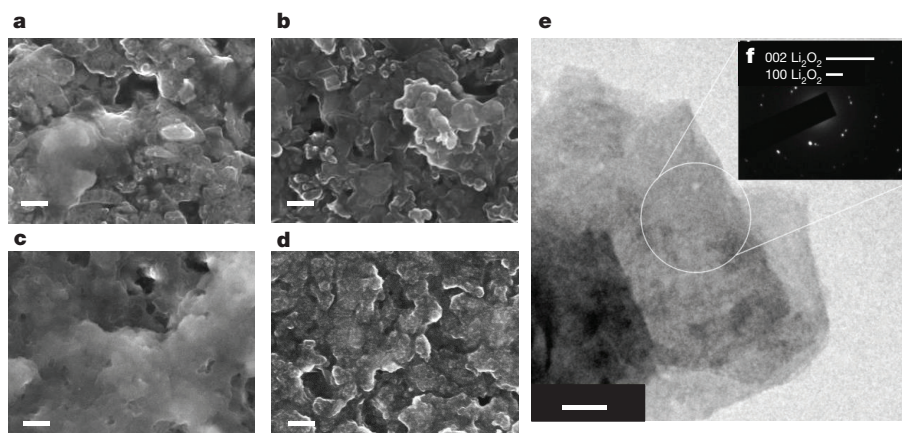


Figure 3 | Electron microscopy of the cathode. **a–d**, The SEM images of lithium–air battery cathodes after the first discharge (**a**), the first charge (**b**), the 250th discharge (**c**) and the 250th charge (**d**). Scale bars, 200 nm.

e, TEM image of a discharged cathode sample. **f**, Diffraction pattern showing crystallinity corresponding to monoclinic lithium peroxide, Li_2O_2 .

after 100 cycles, and for an aged sample (25 cycles, aged for 100 hours; Supplementary Information section 14).

Additional evidence that the presence of CO_2 and H_2O in our lithium–air cell does not cause any detrimental reactions was obtained from the results of a discharge–charge experiment in a pure oxygen environment, using the same cell as for the aforementioned experiments (Supplementary Information section 15). Figure 2g demonstrates similar results for both the lithium–air and lithium–oxygen batteries, indicating that the discharge and charge chemistries are the same in both systems. To examine the stability of the EMIM- BF_4 /DMSO electrolyte in the lithium–air system, we also performed ^1H and ^{13}C nuclear magnetic resonance (NMR) spectroscopic studies on both fresh electrolytes and electrolytes after 550 cycles of operation. The ^1H NMR and ^{13}C NMR spectra (Supplementary Information section 16) show similar peaks for both samples, verifying the stability of the electrolyte during operation of the cell.

The electrochemical reaction during the charge and discharge processes was quantified by *in situ* monitoring of the evolved and consumed gases, respectively, during the first cycle using differential electrochemical mass spectrometry (DEMS)^{2,4,19,20} (Supplementary Information section 17). The DEMS results shown in Fig. 2h indicate an electron/oxygen ratio of 2.07 for the charge reaction and 2.04 for the discharge reaction. These results provide strong evidence for the reversible formation and decomposition of Li_2O_2 as the main product, through a two-electron transfer process.

The morphology and composition of the discharge product was studied using several techniques. SEM images of the cathodes after the 1st and 250th cycles (Fig. 3a–d) indicate a film-like morphology of the discharge products on the cathode surface. This was also observed for samples after 100 cycles (Supplementary Information section 18). X-ray diffraction experiments performed on the discharge product after 550 cycles revealed distinct peaks at 32° and 34° , which were attributed to crystalline (100) and (101) facets of Li_2O_2 (Supplementary Information section 19)^{2,3}. Finally, TEM images and diffraction patterns obtained from the cathode (Fig. 3e, f) show the discharge product to be highly crystalline, with the diffraction pattern corresponding to Li_2O_2 viewed along a *c*-axis direction. Scanning transmission electron microscopy (STEM) imaging was also used to confirm the morphology of the discharge product, showing that Li_2O_2 has a thin film-like structure that consists of extended single crystal domains. EELS spectroscopy was performed on a sample of the discharge product taken from the Li_2O_2 film, and revealed a Li K-edge fine structure consistent with reference Li_2O_2 samples that are stoichiometric and largely free of oxygen vacancies (Supplementary Information section 20).

We carried out density functional theory (DFT) calculations to provide insight into why the cell configuration used in this study can operate in the presence of air. We investigated the Li_2CO_3 anode coating

and its ability to prevent reactions of lithium metal with N_2 and O_2 ; the possible reactions of CO_2 and H_2O with the discharge product at the cathode; and the possible side reactions in the electrolyte with discharge species. DFT calculations concerning the catalytic role of the molybdenum-terminated edge atoms of MoS_2 nanoflakes in the presence of the ionic-liquid electrolyte in a lithium–oxygen battery have been previously reported².

To investigate the role of the anode, DFT calculations (Supplementary Information section 21.1) were carried out on a Li_2CO_3 /Li interface (Supplementary Information section 21.2). The relaxed Li_2CO_3 /Li interface with carbon termination—the most stable termination—is shown in Fig. 4a. Although lithium and Li_2CO_3 are thermodynamically unstable with respect to Li_2O and carbon²¹, the interface appears to be kinetically stable based on the barrier for oxygen migration from a C–O bond to lithium; a barrier of 1.1 eV is found for this reaction pathway (Fig. 4b). In addition, we investigated whether N_2 and O_2 were likely to migrate through Li_2CO_3 by placing them in the (010) channel of Li_2CO_3 (Fig. 4c) and calculating the energy for the relaxed structure relative to the Li_2CO_3 and the free molecules (Supplementary Information section 21.3). The energies were found to be high (1.2–3.2 eV), which indicates that Li_2CO_3 should provide a good barrier to prevent N_2 and O_2 from reaching the lithium anode, assuming no cracks in the structure. Li_2CO_3 should therefore also prevent oxygen crossover—in which oxygen crosses from the cathode to the anode—which is known to result in oxidation of the lithium anode and to limit cycle life⁵. Because H_2O and CO_2 molecules are larger than N_2 and O_2 , their interactions with Li_2CO_3 channels should also be endothermic. On the basis of previous DFT calculations for lithium-ion batteries^{22,23}, the diffusion of lithium through Li_2CO_3 should be quite facile, unlike that of N_2 and O_2 .

The second aspect of our computational studies used DFT calculations to investigate possible detrimental reactions of CO_2 and H_2O with the Li_2O_2 discharge product. We have assumed that the nitrogen will not react with the discharge product, owing to its inertness. The SEM and TEM studies indicated that the Li_2O_2 product has a film-like morphology with few defects. The film-like Li_2O_2 discharge product is consistent with its postulated formation mechanism proposed in a previous study of a MoS_2 cathode material for a lithium–oxygen cell.² DFT calculations (Supplementary Information sections 21.4 and 21.5) indicate considerable binding of a Li_2O_2 molecule and Li_2O_2 cluster to the basal plane of a MoS_2 nanoflake, which could provide sites for the nucleation and growth of Li_2O_2 . The interaction of a Li_2O_2 molecule with the MoS_2 basal plane is shown in Fig. 4d. The high dielectric constant of the ionic liquid/DMSO electrolyte means that Li_2O_2 is sufficiently soluble²⁴ to enable a solution growth mechanism²⁵. Finally, the high discharge rate could contribute to the formation of the thin film found in our cell^{26,27}. This is important because a Li_2O_2 film is expected to have fewer defect sites compared to other morphologies such as

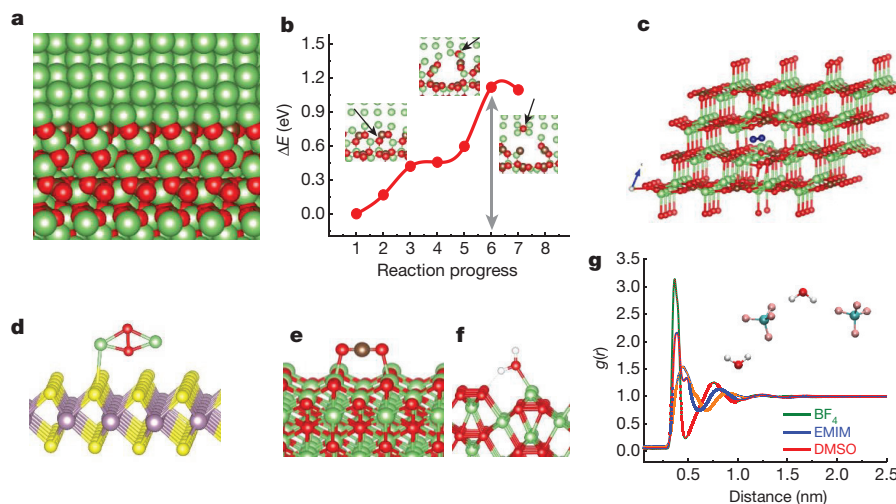


Figure 4 | Computational study of the anode, cathode and electrolyte of the lithium-air battery cell. **a**, Interface between (001) Li_2CO_3 and (100) Li with carbon termination, as calculated by DFT. **b**, DFT profile showing the breaking of the C–O bond of Li_2CO_3 at the interface, and the migration of oxygen to the lithium bulk (black arrows indicate oxygen). **c–f**, Depiction of interactions relevant to the lithium-air battery cell, as calculated by DFT: O_2 in the Li_2CO_3 channel, endothermic by 3.1 eV (**c**);

Li_2O_2 on the MoS_2 -nanoflake basal plane, binding energy 0.83 eV (**d**); CO_2 on the (100) Li_2O_2 surface, binding energy 0.27 eV (**e**); H_2O on the (100) Li_2O_2 surface, binding energy 0.77 eV (**f**). **g**, Classical molecular dynamics radial distribution function showing the interaction of water with BF_4^- , EMIM^+ and DMSO, calculated from the trajectory of a system at 2 mol% water.

nanoparticles or toroids²⁸, thereby reducing decomposition reactions involving CO_2 and H_2O .

To investigate possible reactions of H_2O and CO_2 with the Li_2O_2 surface expected for a thin-film morphology, we carried out DFT calculations of H_2O and CO_2 adsorbed on various relaxed Li_2O_2 surfaces²⁹ in the presence and absence of an electrolyte (Supplementary Information sections 21.6 and 21.7). The most stable structures in the absence of electrolyte are shown in Fig. 4e, f. The strongest adsorption energies for CO_2 and H_2O on the Li_2O_2 surfaces are 0.27 and 0.77 eV, respectively. According to the calculations, neither H_2O nor CO_2 decompose when adsorbed on the Li_2O_2 surfaces. We also included the electrolyte in the calculations in *ab initio* molecular dynamics simulations. The simulations with the electrolyte/ Li_2O_2 interface did not show any specific preference for CO_2 or H_2O to adsorb on the surface or remain in the electrolyte. These results are consistent with calculations of the solvation energies of H_2O and CO_2 in the ionic liquid/DMSO electrolyte, indicating reasonably strong binding in the ionic liquid (Supplementary Information section 21.8). The solvation energies in solution are at least as strong or stronger than the binding to the Li_2O_2 surface. Furthermore, in the presence of the electrolyte, the adsorbed species at the interface do not react with the Li_2O_2 surface.

We also carried out *ab initio* molecular dynamics simulations for CO_2 and H_2O molecules interacting with a defective Li_2O_2 surface containing a single lithium vacancy, in the presence of a mixed ionic liquid/DMSO electrolyte, and found no decomposition such as that which may occur on toroid discharge products (Supplementary Information section 21.7). All of the computational results of the reaction of H_2O and CO_2 with a Li_2O_2 surface are consistent with the lack of evidence for side reactions in the characterization studies.

Finally, we considered side reactions involving the probable discharge species, LiO_2 and Li_2O_2 , with CO_2 or H_2O in the electrolyte using classical molecular dynamics simulations (Supplementary Information section 22.1). The peaks in the radial distribution function in Fig. 4g show the interaction of water molecules with BF_4^- and DMSO from the solvent. Further analysis of hydrogen bonds between these species show that approximately 72% of the water molecules are involved in small clusters of BF_4^- , DMSO and H_2O , all connected through hydrogen bonds (Supplementary Information section 22.2). A representative cluster is shown in the inset of Fig. 4g. Our DFT calculations (Supplementary Information section 22.2) show that the reactions of water with Li_2O_2

or LiO_2 are thermodynamically unfavourable, and reaction with two water molecules would be required. However, the classical molecular dynamics simulations show that, at 2 mol% of water, clusters of two or more water molecules occur only less than 3% of the time. Therefore, there is very little likelihood of the LiO_2 or Li_2O_2 solvated species encountering a water cluster. In the case of CO_2 , the reaction with Li_2O_2 or LiO_2 is thermodynamically favourable but requires reaction with two or more CO_2 molecules; such clusters are highly unlikely (less than 0.2%) based on the classical molecular dynamics simulations.

In summary, the characterization and computational studies have shown that the protected lithium anode, electrolyte blend and high-performance air cathode all work in synergy to provide a lithium-oxygen battery with a long cycle life under simulated air conditions. This new architecture is a promising step towards engineering the next generation of lithium batteries with much higher specific energy density than current lithium-ion batteries.

Online Content Methods, along with any additional Extended Data display items and Source Data, are available in the online version of the paper; references unique to these sections appear only in the online paper.

Received 12 February 2017; accepted 30 January 2018.

- Bruce, P. G., Freunberger, S. A., Hardwick, L. J. & Tarascon, J.-M. Li–O₂ and Li–S batteries with high energy storage. *Nat. Mater.* **11**, 19–29 (2012).
- Asadi, M. *et al.* Cathode based on molybdenum disulfide nanoflakes for lithium–oxygen batteries. *ACS Nano* **10**, 2167–2175 (2016).
- Jung, H.-G., Hassoun, J., Park, J.-B., Sun, Y.-K. & Scrosati, B. An improved high-performance lithium–air battery. *Nat. Chem.* **4**, 579–585 (2012).
- Lu, J. *et al.* A nanostructured cathode architecture for low charge overpotential in lithium–oxygen batteries. *Nat. Commun.* **4**, 2383 (2013).
- Assary, R. S. *et al.* The effect of oxygen crossover on the anode of a Li–O₂ battery using an ether-based solvent: insights from experimental and computational studies. *ChemSusChem* **6**, 51–55 (2013).
- Geng, D. *et al.* From lithium–oxygen to lithium–air batteries: challenges and opportunities. *Adv. Energy Mater.* **6**, 1502164 (2016).
- Christensen, J. *et al.* A critical review of Li/air batteries. *J. Electrochem. Soc.* **159**, R1–R30 (2012).
- García-Araez, N. & Novák, P. Critical aspects in the development of lithium–air batteries. *J. Solid State Electrochem.* **17**, 1793–1807 (2013).
- Zhang, T. & Zhou, H. A reversible long-life lithium–air battery in ambient air. *Nat. Commun.* **4**, 1817 (2013).
- Das, S. K., Chai, J., Rahman, S. & Sarkar, A. Synthesis, characterization and performance evaluation of an advanced solid electrolyte and air cathode for rechargeable lithium–air batteries. *J. Mater. Sci. Chem. Eng.* **4**, 74–89 (2016).

11. Wu, S. *et al.* A synergistic system for lithium–oxygen batteries in humid atmosphere integrating a composite cathode and a hydrophobic ionic liquid-based electrolyte. *Adv. Funct. Mater.* **26**, 3291–3298 (2016).
12. Gallagher, K. G. *et al.* Quantifying the promise of lithium–air batteries for electric vehicles. *Energy Environ. Sci.* **7**, 1555–1563 (2014).
13. Pasierb, P., Komornicki, S., Rokita, M. & Rękas, M. Structural properties of Li_2CO_3 – BaCO_3 system derived from IR and Raman spectroscopy. *J. Mol. Struct.* **596**, 151–156 (2001).
14. Bi, Y. *et al.* Stability of Li_2CO_3 in cathode of lithium ion battery and its influence on electrochemical performance. *RSC Advances* **6**, 19233–19237 (2016).
15. Sina, M. *et al.* Investigation of SEI layer formation in conversion iron fluoride cathodes by combined STEM/EELS and XPS. *J. Phys. Chem. C* **119**, 9762–9773 (2015).
16. Imanishi, N., Luntz, A. C. & Bruce, P. (eds) *The Lithium Air Battery: Fundamentals* (Springer, 2014).
17. Du, P. *et al.* Compatibility of lithium salts with solvent of the non-aqueous electrolyte in Li-O_2 batteries. *Phys. Chem. Chem. Phys.* **15**, 5572–5581 (2013).
18. Lu, Y.-C. *et al.* In situ ambient pressure X-ray photoelectron spectroscopy studies of lithium–oxygen redox reactions. *Sci. Rep.* **2**, 715 (2012).
19. McCloskey, B. D., Bethune, D. S., Shelby, R. M., Girishkumar, G. & Luntz, A. C. Solvents' critical role in nonaqueous lithium–oxygen battery. *J. Phys. Chem. Lett.* **2**, 1161–1166 (2011).
20. Lu, J. *et al.* A lithium–oxygen battery based on lithium superoxide. *Nature* **529**, 377–382 (2016).
21. Leung, K., Soto, F., Hankins, K., Balbuena, P. B. & Harrison, K. L. Stability of solid electrolyte interphase components on lithium metal and reactive anode material surfaces. *J. Phys. Chem. C* **120**, 6302–6313 (2016).
22. Iddir, H. & Curtiss, L. A. Li ion diffusion mechanisms in bulk monoclinic Li_2CO_3 crystals from density functional. *J. Phys. Chem. C* **114**, 20903–20906 (2010).
23. Shi, S., Qi, Y., Li, H. & Hector, L. G. Defect thermodynamics and diffusion mechanisms in Li_2CO_3 and implications for the solid electrolyte interphase in Li-ion batteries. *J. Phys. Chem. C* **117**, 8579–8593 (2013).
24. Cheng, L. *et al.* Computational studies of solubilities of LiO_2 and Li_2O_2 in aprotic solvents. *J. Electrochem. Soc.* **164**, E3696–E3701 (2017).
25. Johnson, L. *et al.* The role of LiO_2 solubility in O_2 reduction in aprotic solvents and its consequences for Li– O_2 batteries. *Nat. Chem.* **6**, 1091–1099 (2014).
26. Welland, M. J. *et al.* An atomistically informed mesoscale model for growth and coarsening during discharge in lithium–oxygen batteries. *J. Chem. Phys.* **143**, 224113 (2015).
27. Adams, B. D. *et al.* Current density dependence of peroxide formation in the Li– O_2 battery and its effect on charge. *Energy Environ. Sci.* **6**, 1772 (2013).
28. Xiao, D. *et al.* Direct observation of ordered oxygen defects on the atomic scale in Li_2O_2 for Li– O_2 batteries. *Adv. Energy Mater.* **5**, 1400664 (2015).
29. Lau, K. C., Curtiss, L. A. & Greeley, J. Density functional investigation of the thermodynamic stability of lithium oxide bulk crystalline structures as a function of oxygen pressure. *J. Phys. Chem. C* **115**, 23625–23633 (2011).

Supplementary Information is available in the online version of the paper.

Acknowledgements The work of A.S.-K., M. A., B. S. and P. A. was supported by the National Science Foundation (NSF-DMREF Award #1729420). Work

by B.N., K.C.L., R.S.A., L.A.C., R.F.K., A.M., X. H. and J.R.J. was supported by the Joint Center for Energy Storage Research (JCESR), an Energy Innovation Hub funded by the US Department of Energy, Office of Science, Basic Energy Sciences. Work by A.N. and C.L. was supported by the Center for Electrical Energy Storage: Tailored Interfaces, an Energy Frontier Research Center funded by the US Department of Energy, Office of Science, Office of Basic Energy Sciences. The work of F.K.A. and K.K. was supported by a University of Illinois at Chicago start-up fund. C.L. was also supported by programme development funds provided by the Chemical Sciences and Engineering division at Argonne National Laboratory. We acknowledge the MRSEC Materials Preparation and Measurement Laboratory shared user facility at the University of Chicago (NSFDMR-1420709); the EPIC facility (NUANCE Center, Northwestern University), which has received support from the MRSEC program (NSF DMR-1121262) at the Materials Research Center; the Nanoscale Science and Engineering Center (NSF EEC-0647560) at the International Institute for Nanotechnology; and the State of Illinois, through the International Institute for Nanotechnology. This work also made use of the Integrated Molecular Structure Education and Research Center at Northwestern University, which has received support from the Soft and Hybrid Nanotechnology Experimental (SHyNE) Resource (NSF NNCI-1542205). The acquisition of the UIC JEOL JEM ARM200CF was supported by an MRI-R² grant from the National Science Foundation (DMR-0959470). The use of instrumentation at University of Illinois at Chicago Research Resources Center (RRC-East) is acknowledged. A. Nicholls at UIC's Electron Microscopy Service is also acknowledged for help and support. This research used high performance computing resources of the Argonne Leadership Computing Facility, which is a DOE Office of Science User Facility supported under Contract DE-AC02-06CH11357. Use of the Center for Nanoscale Materials, an Office of Science user facility, was supported by the US Department of Energy, Office of Science, Office of Basic Energy Sciences, under contract DE-AC02-06CH11357. F.K.A. acknowledges the use of the University of Illinois at Chicago High Performance Computing Cluster to perform molecular dynamics simulations. We thank K. Gallagher, P. Redfern, H.-H. Wang, J. Jureller and X. Chen.

Author Contributions A.S.-K. and M.A. conceived the idea. M.A., B.S., P.A. and M.G. performed the electrochemical experiments. M.A. and B.S. synthesized the MoS_2 nanoflakes. M.A., B.S., P.A. and P.Y. carried out characterization. A.S.-K. supervised the electrochemical experiments. B.N., K.C.L., R.S.A. and L.A.C. carried out the computational studies of electrolytes. C.L. and A.T.N. performed computational studies of surfaces and the Li_2CO_3 coating. J.R.J., X.H., A.M. and R.F.K. carried out STEM and EELS experiments. K.K. and F.K.-A. performed classical molecular dynamics simulations. All of the authors contributed to the manuscript before submission.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare competing financial interests: details are available in the online version of the paper. Readers are welcome to comment on the online version of the paper. Publisher's note: Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations. Correspondence and requests for materials should be addressed to L.A.C. (curtiss@anl.gov) or A.S.-K. (salehikh@uic.edu).

Reviewer Information *Nature* thanks S. Soon, G. Yu and the other anonymous reviewer(s) for their contribution to the peer review of this work.

METHODS

Anode preparation. The battery cell was assembled using commercial pure lithium chips (>99.9%) with a thickness of 0.25 mm as the anode, and MoS₂ nanoflakes coated on a gas diffusion layer as the cathode. The electrolyte of the cell was composed of 25 vol% 1-ethyl-3-methylimidazolium tetrafluoroborate (EMIM-BF₄) (HPLC grade, >99.0%) and 75 vol% dimethyl sulfoxide (DMSO) with 0.1 M lithium bis(trifluoromethanesulfonyl)imide (LiTFSI) (>99.0%) as a lithium salt. The cell was run for ten continuous cycles, each cycle consisting of a one-hour charge process followed by a one-hour discharge process, in an environment of pure CO₂.

Cathode preparation. The MoS₂ nanoflakes were synthesized using a liquid exfoliation method², in which 300 mg of MoS₂ powder (99%) was dispersed in 60 ml isopropyl alcohol (>99.5%). The solution was exfoliated for 20 h and centrifuged for 1 h to extract the supernatant (the top two-thirds of the centrifuged solution). MoS₂ nanoflakes were then coated on a conductive substrate of the gas diffusion layer (0.2 mm thickness, 80% porosity) to reach a catalyst loading of 0.1 mg cm⁻². Prepared cathodes were dried in a vacuum oven for 24 h at 80 °C to stabilize the cathode and to remove impurities.

Electrochemical characterization. Electrochemical impedance spectroscopy experiments were performed using coin cells under identical experimental conditions of 700 mV overpotential and a frequency range of 10 Hz to 100 kHz.

For discharge-charge experiments, the lithium-air Swagelok battery set-up consisting of MoS₂ nanoflakes as the cathode, 0.1 M LiTFSI as the lithium salt in EMIM-BF₄/DMSO (25/75 volume ratio) as the electrolyte, and protected lithium as the anode was operated with a constant applied current density of 500 mA g⁻¹.

Characterization techniques. Raman spectroscopy experiments were performed using a Horiba LabRAM HR Evolution confocal Raman microscope. The samples were sealed between two transparent glasses in an argon-filled glove box. The instrument was configured with a 785-nm laser source, 1,200 g mm⁻¹ grating, a Horiba Andor detector and a LCD objective with a modifiable optical ring that enables aberration correction according to glass thickness.

X-ray photoelectron spectroscopy experiments were performed on a Thermo Scientific ESCALAB 250Xi instrument. The instrument was equipped with an electron flood and scanning ion gun. To prevent oxidation and any contamination, a mobile glove box filled with argon was used for transferring the samples into the loading chamber of the instrument.

The DEMS experiment was carried out in a custom-made Swagelok battery set-up. The experimental set-up consisted of a potentiostat (MTI) and a mass spectrometer (Hiden Analytical) operating under ultra-high vacuum. The DEMS was calibrated by injecting into the mass spectrometer standard samples of pure O₂ (99.99%, Praxair) in research-grade argon (99.99%, Praxair) at known concentrations, and measuring the corresponding partial pressures of O₂.

STEM measurements were carried out on an aberration-corrected JOEL JEM-ARM200CF instrument equipped with a cold field-emission electron source and a post-column Gatan Enfina EELS spectrometer. An acceleration voltage of 80 kV was used for both imaging and EELS to reduce beam-induced damage and contamination.

Theoretical methods. Periodic calculations of the surface interactions and Li₂CO₃/Li interfaces were carried out with the Vienna *ab initio* simulation package (VASP)³⁰ code with plane wave basis sets and projector-augmented wave pseudopotentials³¹. The exchange-correlation functional was treated within the

generalized gradient approximation of Perdew-Burke-Ernzerhof³². *Ab initio* molecular dynamics simulations of the electrolyte were performed within the generalized gradient approximation using plane-wave basis sets as implemented in VASP³⁰. Similar to the static calculations, we described the exchange correlation via Perdew-Burke-Ernzerhof functionals³² and use projector-augmented wave pseudopotentials³¹ supplied by VASP for the *ab initio* molecular dynamics simulations. For details of specific calculations, see Supplementary Information section 21.

Quantum chemical calculations were carried out to investigate the thermodynamic stability of solvated structures of CO₂ and H₂O using the Gaussian09 code³³, using the 6-31+G(d,p) basis set with the B3LYP functional.

Classical molecular dynamics simulations were carried out using the software package GROMACS 5^{34,35} with the GROMOS 53A6 force field³⁶. To represent atomic interactions of water molecules, we used the SPC³⁷ as well as the SPC/E³⁸ water models owing to their considerably different self-diffusion coefficients^{38–40}. All simulations were carried out using isotropic constant pressure control using the Berendsen Barostat⁴¹, with a coupling constant of 1.0 ps and a compressibility of 2.755×10^{-5} bar⁻¹ used for all equilibrium and production simulations. Temperature was maintained at 300 K using a Berendsen Thermostat⁴¹, with a coupling constant of 0.1 ps and a time step of 1 fs. The Verlet cut-off scheme was used. Coulomb interactions were calculated using the particle mesh Ewald method with a Coulomb radius set to 1.4 nm. Van der Waals forces were evaluated using a switching function from 0.8 to 1.4 nm.

Data availability. The data that support the findings of this study are available from the corresponding authors upon reasonable request.

30. Kresse, G. & Furthmüller, J. Efficiency of *ab-initio* total energy calculations for metals and semiconductors using a plane-wave basis set. *Comput. Mater. Sci.* **6**, 15–50 (1996).
31. Kresse, G. From ultrasoft pseudopotentials to the projector augmented-wave method. *Phys. Rev. B* **59**, 1758–1775 (1999).
32. Perdew, J. P., Burke, K. & Ernzerhof, M. Generalized gradient approximation made simple. *Phys. Rev. Lett.* **77**, 3865–3868 (1996).
33. Frisch, M. J. *et al.* Gaussian 09, Revision D.01. (Gaussian Inc., 2009).
34. Berendsen, H. J. C., van der Spoel, D. & van Drunen, R. GROMACS: A message-passing parallel molecular dynamics implementation. *Comput. Phys. Commun.* **91**, 43–56 (1995).
35. Abraham, M. J. *et al.* GROMACS: High performance molecular simulations through multi-level parallelism from laptops to supercomputers. *SoftwareX* **1–2**, 19–25 (2015).
36. Horta, B. A. C. *et al.* A GROMOS-compatible force field for small organic molecules in the condensed phase: the 2016H66 parameter set. *J. Chem. Theory Comput.* **12**, 3825–3850 (2016).
37. Berendsen, H. J. C. *et al.* in *Intermolecular Forces* Vol. 14 (ed. Pullman, B.) 331–342 (Springer, 1981).
38. Berendsen, H. J. C., Grigera, J. R. & Straatsma, T. P. The missing term in effective pair potentials. *J. Phys. Chem.* **91**, 6269–6271 (1987).
39. Mark, P. & Nilsson, L. Structure and dynamics of liquid water with different long-range interaction truncation and temperature control methods in molecular dynamics simulations. *J. Comput. Chem.* **23**, 1211–1219 (2002).
40. Mark, P. & Nilsson, L. Structure and dynamics of the TIP3P, SPC, and SPC/E water models at 298 K. *J. Phys. Chem. A* **105**, 9954–9960 (2001).
41. Berendsen, H. J. C., Postma, J. P. M., van Gunsteren, W. F., DiNola, A. & Haak, J. R. Molecular dynamics with coupling to an external bath. *J. Chem. Phys.* **81**, 3684–3690 (1984).

Isotopic evolution of the protoplanetary disk and the building blocks of Earth and the Moon

Martin Schiller¹, Martin Bizzarro¹ & Vera Assis Fernandes^{2,3}

Nucleosynthetic isotope variability among Solar System objects is often used to probe the genetic relationship between meteorite groups and the rocky planets (Mercury, Venus, Earth and Mars), which, in turn, may provide insights into the building blocks of the Earth–Moon system^{1–5}. Using this approach, it has been inferred that no primitive meteorite matches the terrestrial composition and the protoplanetary disk material from which Earth and the Moon accreted is therefore largely unconstrained⁶. This conclusion, however, is based on the assumption that the observed nucleosynthetic variability of inner-Solar-System objects predominantly reflects spatial heterogeneity. Here we use the isotopic composition of the refractory element calcium to show that the nucleosynthetic variability in the inner Solar System primarily reflects a rapid change in the mass-independent calcium isotope composition of protoplanetary disk solids associated with early mass accretion to the proto-Sun. We measure the mass-independent $^{48}\text{Ca}/^{44}\text{Ca}$ ratios of samples originating from the parent bodies of ureilite and angrite meteorites, as well as from Vesta, Mars and Earth, and find that they are positively correlated with the masses of their parent asteroids and planets, which are a proxy of their accretion timescales. This correlation implies a secular evolution of the bulk calcium isotope composition of the protoplanetary disk in the terrestrial planet-forming region. Individual chondrules from ordinary chondrites formed within one million years of the collapse of the proto-Sun⁷ reveal the full range of inner-Solar-System mass-independent $^{48}\text{Ca}/^{44}\text{Ca}$ ratios, indicating a rapid change in the composition of the material of the protoplanetary disk. We infer that this secular evolution reflects admixing of pristine outer-Solar-System material into the thermally processed inner protoplanetary disk associated with the accretion of mass to the proto-Sun. The identical calcium isotope composition of Earth and the Moon reported here is a prediction of our model if the Moon-forming impact involved protoplanets or precursors that completed their accretion near the end of the protoplanetary disk's lifetime.

A protoplanetary disk forms as a result of mass accretion from the collapse of the envelope onto the star. The rate of mass accretion onto the star via the disk is typically high in the early stages, resulting in elevated temperatures within a few astronomical units of the star⁸. In the Solar System, mass accretion to the proto-Sun resulted in thermal processing of pristine infalling molecular-cloud material, including the selective destruction of presolar carriers of nucleosynthetic isotope anomalies. This process is revealed by the correlated variability in the abundance of nuclides of distinct nucleosynthetic origin, such as ^{46}Ti – ^{50}Ti and ^{43}Ca – ^{46}Ca – ^{48}Ca , among Solar System reservoirs^{2,3}. Parent bodies of meteorites formed from inner, thermally processed disk material show depletions in these isotopes relative to bodies formed in the outer-disk regions. The mass-independent isotopic compositions of various elements can therefore characterize the source of the precursor material to the rocky planets. However, this requires an understanding of the secular evolution of

the nucleosynthetic composition of the disk material that accreted to growing planetary bodies. Unlike nucleosynthetic tracers (such as Ti and Cr, which are present in planets as trace elements) and siderophile elements (such as Mo and Ru, which can help identify the source of the metal fraction of a body), calcium is one of the main constituents of rock-forming minerals and hence provides robust constraints on the precursor material to the rocky planets. To probe the Ca isotope evolution of the protoplanetary disk, we determined the mass-independent ^{48}Ca composition of selected Solar System objects, including ordinary and carbonaceous chondrites, ureilites, eucrites, angrites, martian and lunar meteorites, as well as individual chondrules from ordinary and carbonaceous chondrites (Fig. 1). We report the mass-independent Ca data as $\mu^{48}\text{Ca}$ values, which express the calcium isotope composition of the sample relative to that of the SRM 915b standard: $\mu^{48}\text{Ca} = [(^{48}\text{Ca}/^{44}\text{Ca})_{\text{sample}} / (^{48}\text{Ca}/^{44}\text{Ca})_{\text{SRM915b}} - 1] \times 10^6$.

The growth of asteroids and Mars-sized embryos can result from the gas-drag-assisted accretion of millimetre-sized chondrules within about

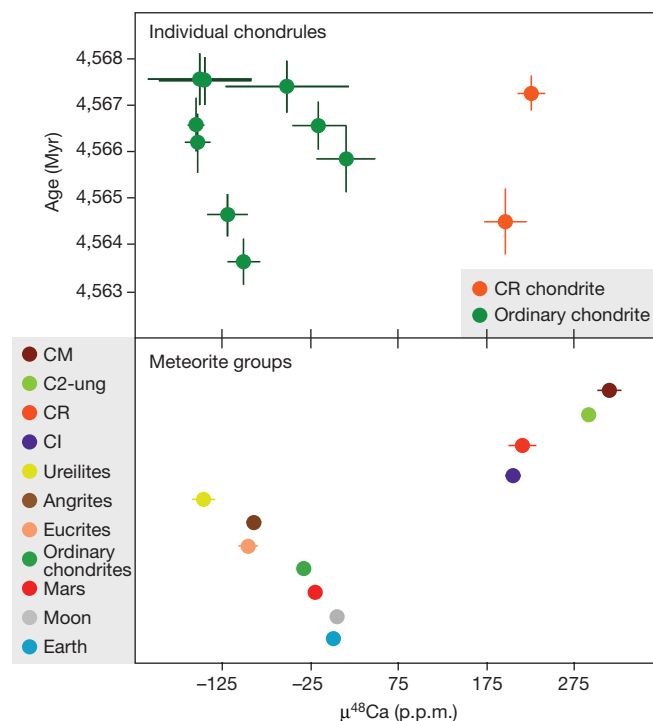


Figure 1 | Mass-independent $\mu^{48}\text{Ca}$ data for individual chondrules and bulk meteorites. Chondrule data are shown versus their age⁷, as determined by Pb–Pb dating. Bulk rock $\mu^{48}\text{Ca}$ data are shown as a Caltech plot. Uncertainties are two standard deviations for Pb–Pb ages and two standard errors of the mean for $\mu^{48}\text{Ca}$ values. CM, Mighei-type chondrites; C2-ung, ungrouped carbonaceous chondrite of petrological type 2.

¹Centre for Star and Planet Formation and Natural History Museum of Denmark, University of Copenhagen, Øster Voldgade 5–7, DK-1350, Denmark. ²Museum für Naturkunde, Leibniz-Institut für Evolutions- und Biodiversitätsforschung, Berlin 10115, Germany. ³Instituto Dom Luiz, Faculdade de Ciências, Universidade de Lisboa, 1749-016 Lisboa, Portugal.

3 million years (Myr)^{9,10}. The ages obtained by U-corrected Pb–Pb dating and the initial Pb isotope compositions of the chondrules studied indicate a primary formation age restricted to the first million years of disk evolution, although some objects were remelted at later times⁷. Thus, these chondrules provide insights into the spatial and temporal evolution of the $\mu^{48}\text{Ca}$ value of disk solids at early times (Fig. 1). The CR chondrite (Renazzo-type) chondrules have calcium isotope compositions comparable to those of CI carbonaceous (Ivuna-type) chondrites, indicating that their precursor material escaped extensive thermal processing. By contrast, ordinary chondrite chondrules show variable $\mu^{48}\text{Ca}$ values, ranging from terrestrial values to $\mu^{48}\text{Ca}$ deficits of -161 ± 27 p.p.m., where the uncertainties represent twice the standard error of the mean. Our bulk meteorite data are consistent with the conclusions of earlier studies^{3,11}, namely, that bodies formed in the inner disk show systematic depletions in ^{48}Ca , whereas carbonaceous outer-Solar-System bodies exhibit ^{48}Ca excesses relative to Earth (Fig. 1).

Given the analytical uncertainty in the ages of individual chondrules, it is difficult to distinguish whether their ^{48}Ca deficits relative to the solar composition reflect spatial heterogeneity or temporal evolution of the bulk composition of inner-disk solids. Astronomical observations of young stars and their disks indicate that the main accretion epoch of thermally unprocessed envelope material occurs over timescales comparable to those of primary chondrule formation, namely, during the first million years of disk evolution¹². Assuming that CI chondrites have approximately the average composition of the envelope material, admixing this material into the inner disk will result in a progressive increase in the $\mu^{48}\text{Ca}$ value of solids and bodies formed during this epoch. To test this hypothesis, we turn to the calcium isotope composition of differentiated asteroids and planetary bodies because their $\mu^{48}\text{Ca}$ values reflect the average compositions of their precursors throughout the accretion history of these bodies.

Recent models^{9,10} of the formation and growth of asteroidal bodies and planetary embryos suggest a two-stage process, where first-generation bodies with characteristic sizes of about 100 km form rapidly by streaming instabilities, followed by continuous growth dominated by gas-drag-assisted accretion of millimetre-sized particles for bodies with radii larger than about 200 km. This results in the formation of Mars-sized embryos over typical disk lifetimes of less than 5 Myr. Importantly, formation of the first embryos leads to the excitation of the inclinations of the smaller asteroids, which disconnects the asteroids from the chondrules in the mid-plane layer and hence terminates their accretion. Assuming a similar rate of accretion in the terrestrial-planet-forming region, these models predict that the final mass of a rocky body is a function of its accretion timescale. Although the masses of Earth, Mars and Vesta (the parent body of howardite, eucrite and diogenite meteorites¹³) can be inferred, the parent bodies of ureilite and angrite meteorites are not known. However, the presence of a core dynamo in the angrite parent body¹⁴, magmatic differentiation, and volcanic evolution indicate a similar thermal history to Vesta^{15,16}, and the retention of basaltic lavas on its surface suggests that the angrite parent body may have been comparable in size to Vesta. The mass of the ureilite parent body was calculated using a density of 3.22 g cm^{-3} and a radius of $105 \pm 25 \text{ km}$ (refs 17 and 18). We note that the absence of basaltic meteorites from the ureilite parent body, which could indicate loss of basaltic lavas through explosive volcanism, is consistent with a radius no larger than 100 km (ref. 19). Figure 2a shows that the ^{48}Ca composition of differentiated bodies in the inner Solar System systematically increases with mass, indicating a secular evolution of the time-integrated average ^{48}Ca composition of the disk material that accreted to form the terrestrial planets. Thus, the inner-disk ^{48}Ca nucleosynthetic variations predominantly reflect progressive admixing of pristine matter from the outer Solar System into an initially thermally processed dust reservoir during the early stages of disk evolution. This establishes that the nucleosynthetic variability recorded by inner-disk bodies is primarily controlled by the formation timescales of their

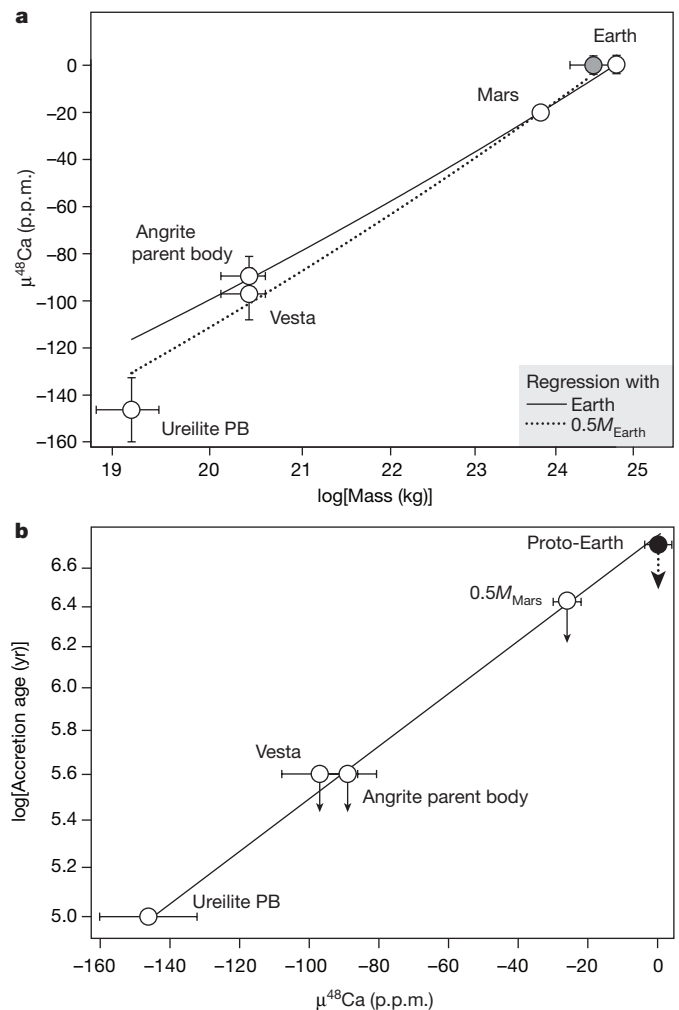


Figure 2 | The $\mu^{48}\text{Ca}$ value of planetary bodies versus their mass and their maximum accretion age. **a**, Comparison between $\mu^{48}\text{Ca}$ values of the ureilite parent body, the angrite parent body, Vesta, Mars and Earth with their respective masses on a logarithmic scale. The solid line is a linear least-squares fit ($y = a + b \times x$), where $a = -517.08 \pm 0.11$ and $b = 20.878 \pm 0.005$, and the uncertainties correspond to one standard deviation. We assigned 50% uncertainty to the masses of the angrite parent body and the eucrite parent body (Vesta). The dotted line shows a regression through the data assuming a precursor with mass $(0.50 \pm 0.25)M_{\text{Earth}}$, where M_{Earth} is Earth's mass. Regressions to data obtained using alternative masses for the ureilite parent body and Earth (precursors) are shown in Extended Data Fig. 1. **b**, Accretion ages of the ureilite parent body, the angrite parent body, Vesta, the proto-Earth, and Mars at half of its current mass (see Methods and refs 20, 21 and 23), shown in a logarithmic scale with respect to the formation of the Solar System, versus the $\mu^{48}\text{Ca}$ values of their parent bodies. Downward arrows indicate that the accretion ages shown are upper limits. A linear regression with $a = 6.7232$ and $b = 0.011886$ is used to calculate the accretion age of the proto-Earth at 5.3 Myr on the basis of its $\mu^{48}\text{Ca}$ value. The accretion age of the proto-Earth probably reflects the accretion age of its precursors, which may include a number of planetary embryos. Thus, Earth's final accretion occurred after disk dissipation from colliding embryos. The formation of a 0.9-Earth-mass object at 1 AU from the Sun within 5 Myr is possible via pebble accretion¹⁰, although early accretion of the proto-Earth is inconsistent with the present-day W isotope composition of Earth's mantle³³. However, a better understanding of the W isotope evolution of the proto-Earth in the framework of pebble accretion and a thorough assessment of the extent of metal–silicate fractionation during the Moon-forming impact are required to evaluate this model.

precursors. This interpretation is generally consistent with data from other nucleosynthetic tracers of lithophile affinity when considering the uncertainties of the data (see Methods).

The inner-Solar-System chondrules investigated here suggest that the $\mu^{48}\text{Ca}$ value of disk solids evolved from an ureilite-like to a terrestrial composition within about 1 Myr of the formation of the Solar System. Although these data indicate a rapid isotopic evolution of accreting disk solids, they provide no constraints on the rate of change of the average bulk composition of the inner disk. However, this information can be obtained from the $\mu^{48}\text{Ca}$ values of differentiated asteroids and planetary bodies because their composition reflects the time-integrated average of the disk composition at the time of their accretion. Two ordinary chondrite chondrules that have U-corrected Pb–Pb ages within about 100,000 years of the formation of the Solar System have $\mu^{48}\text{Ca}$ values identical to that of the ureilite parent body—the differentiated asteroid with the largest $\mu^{48}\text{Ca}$ deficit. Short-lived-radiionuclide chronology and thermal modelling indicate that the ureilite parent body completed its accretion less than 100,000 years from the formation of the Solar System (see Methods). Collectively, these observations suggest that the ureilite $\mu^{48}\text{Ca}$ value of -146 ± 14 p.p.m. represents an inner-disk dust reservoir initially depleted in ^{48}Ca through thermal processing. When Vesta and the angrite parent bodies accreted, less than 0.25 ± 0.15 Myr from the formation of the Solar System^{20,21}, the time-integrated average $\mu^{48}\text{Ca}$ of inner-disk rocky bodies had evolved to -93.2 ± 7.1 p.p.m., and it reached -20.0 ± 2.8 p.p.m. and 0.2 ± 3.9 p.p.m., respectively, when Mars and the precursor(s) of Earth completed their accretion.

It has been established that the protoplanetary disk was dissipated at the time of the formation of the impact-generated CB (Bencubbin-type) Gujba chondrite chondrules, $4,562.49 \pm 0.21$ million years ago²², implying cessation of the infall of outer-Solar-System material to the inner disk. Thus, the timing of disk dissipation marks a transition point, after which the $\mu^{48}\text{Ca}$ value of inner-disk bodies can no longer be modified by infall. Given the observed relationship between $\mu^{48}\text{Ca}$ values and planetary masses (Fig. 2a), the timing of the accretion of Earth's precursor can be deduced, assuming that the accretion timescales of Mars, Vesta, and the angrite and ureilite parent bodies are well understood. Hf–W systematics of martian meteorites²³ suggest that Mars reached half of its mass $1.8^{+0.9}_{-1.0}$ Myr after the formation of the Solar System, whereas thermal modelling and ^{26}Al – ^{26}Mg data constrain the accretion of Vesta and the angrite parent body to less than 0.4 Myr^{20,21}. Accretion of the ureilite parent body is believed to have been completed by about 0.1 Myr after the formation of the Solar System to ensure partial differentiation from the decay of ^{26}Al (see Methods). Using these estimates, we show in Fig. 2b that accretion of the proto-Earth or its precursors, which may include a number of planetary embryos (see Fig. 2 legend), was completed within about 5.3 Myr. These observations suggest that the $\mu^{48}\text{Ca}$ value of Earth represents the average bulk inner-disk composition before its dissipation. The addition rate of material from the outer Solar System to the inner disk can be evaluated from the secular change in the average inner-disk ^{48}Ca composition, which is inferred from differentiated asteroids and planets. Accepting an ureilite-like $\mu^{48}\text{Ca}$ value for the bulk inner disk at 0.1 Myr, we show in Fig. 3a that the rate of addition of material from the outer Solar System to the inner disk mirrors the typical mass accretion rates of low-mass protostars²⁴. Therefore, this flux of material appears to be associated with the infall of envelope material and hence the growth of the proto-Sun.

Our model for the secular $\mu^{48}\text{Ca}$ evolution of the inner protoplanetary disk provides insights into the nature of the precursor material to planetesimals and planets. The compositions of Vesta and the angrite parent body require mixing of 15% of thermally unprocessed CI-like material with their precursors. For Mars and the proto-Earth, the amount of thermally unprocessed material needed to account for their bulk $\mu^{48}\text{Ca}$ values is 36% and 42%, respectively. If correct, these evaluations imply that more than half of the mass of the inner disk was already locked into asteroidal bodies or planetary embryos about 100,000 years after the collapse of the proto-Sun (Fig. 3b). Thus, the $\mu^{48}\text{Ca}$ values of terrestrial planets reflect their accretion histories throughout the lifetime of the disk. By contrast, most chondrites represent fragments of bodies formed during the late stages of accretion and, as such, can

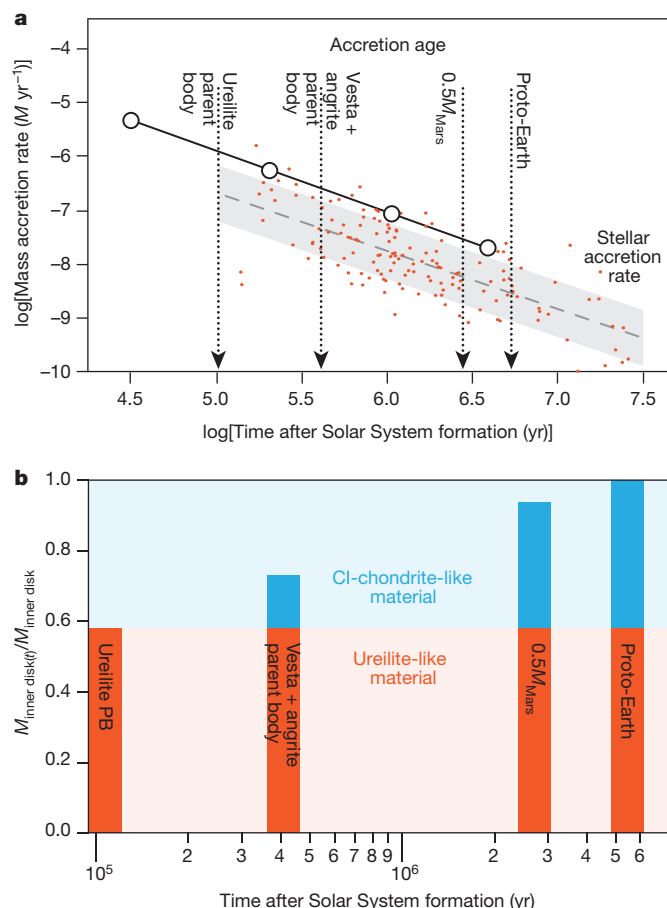


Figure 3 | Accretion rate of inner-disk mass versus time. **a**, Mass accretion rates (in inner-disk masses, M , per year) of the inner protoplanetary disk, calculated from the $\mu^{48}\text{Ca}$ values and maximum accretion timescales of planetary bodies (white circles), assuming that the $\mu^{48}\text{Ca}$ signature of a body is representative of the entire inner-disk mass at that time. Therefore, the accretion rate of the inner protoplanetary disk reflects the amount of mass locked inside bodies at any given time relative to the final disk mass, which is taken as the mass of asteroids, moons and planets located sunwards of Jupiter (see Methods). Given the apparent power-law decline in the accretion rates, we show each calculated value at the logarithmic centre of the corresponding time interval, $\Delta \log(t)/2$, where t is the time after the formation of the Solar System. For comparison, we also show observed stellar accretion rates (in solar masses, M_{\odot} , per year) for $0.3M_{\odot}$ – M_{\odot} objects (red dots) scaled to a $0.7M_{\odot}$ star, as well as the best linear fit (dashed line) and one standard deviation (shaded area) of these data²⁴. **b**, Proportion of mass locked inside asteroids, embryos and planets at their maximum accretion age ($M_{\text{inner disk}}(t)/M_{\text{inner disk}}$), assuming that the composition of each parent body is representative of the bulk disk at the time of accretion. The fractions are calculated as mixtures between a ureilite-like $\mu^{48}\text{Ca}$ value and outer-Solar-System dust, represented by CI chondrites (see Methods).

at best represent only a snapshot of the disk composition at the time of their accretion. However, given the small size of their parent bodies, it is unclear whether their compositions are representative of the average accreting material at the time of their formation, which can be also influenced by variable proportions of chondrule and matrix. For example, the ordinary chondrite chondrules analysed here have a weighted average $\mu^{48}\text{Ca}$ of -129 ± 50 p.p.m., making the bulk rock composition very sensitive to the chondrule-to-matrix ratio. Thus, chondrites cannot be used to assess the nature of the precursor material to rocky planets reliably.

To explain the observed nucleosynthetic dichotomy between inner- and outer-Solar-System bodies, it has been proposed²⁵ that the transport of material from the outer Solar System to the inner disk

was stopped by the opening of a disk gap related to the formation of Jupiter's core, about 1 Myr after the formation of the Sun. However, numerical simulations and astronomical observations suggest that such an opening would not quench mass accretion to the protostar, but would limit the inward transport of large dust grains by filtration^{26,27}. Coagulation of smaller dust particles Sunwards of the disk gap would continue to fuel planetary growth by pebble accretion. The apparent lack of this material in ordinary chondrites may reflect the fact that the main growth mechanism of small bodies—namely, the streaming instability—is less efficient in accreting smaller particles. Testing this hypothesis, however, requires numerical simulations that better describe the relationship between size sorting during planetesimal formation by the streaming instability and size sorting by pebble accretion. Lastly, our results and interpretations imply that more than half of the mass of the inner disk was already locked into sizeable bodies with a $\mu^{48}\text{Ca}$ values similar to that of ureilites about 100,000 years after the collapse of the proto-Sun (Fig. 3b). Taking into account mass balance, mixing of pristine CI-like dust from the outer Solar System with this inner-disk reservoir can at most raise the bulk $\mu^{48}\text{Ca}$ value of the disk to approximately the terrestrial one, thereby preserving an isotopic contrast between rocky planets and carbonaceous asteroidal bodies. Therefore, we infer that the nucleosynthetic dichotomy between inner- and outer-Solar-System bodies reflects the rapid accretion of thermally processed disk material into asteroidal bodies or planetary embryos and does not require the early formation of a disk gap.

The giant impact theory for the formation of the Earth–Moon system proposes that a Mars-sized embryo collided with the proto-Earth and ejected material into an Earth-orbiting disk, which subsequently accumulated into the Moon²⁸. Because models predict that most of the Moon's mass is derived from the impactor, this theory is difficult to reconcile with the nearly identical isotopic compositions of Earth and the Moon for lithophile elements^{1,5,29}. Therefore, a number of alternative models have been proposed that allow for disk–planet compositional equilibration^{30–32}. Our high-precision data define a $\mu^{48}\text{Ca}$ value of 3.7 ± 1.9 p.p.m. for the Moon. In the framework of the canonical model²⁹, this requires the $\mu^{48}\text{Ca}$ values of the proto-Earth and the impactor to be within 10 p.p.m. of each other. Our results suggest that the bulk $\mu^{48}\text{Ca}$ composition of the inner protoplanetary disk evolved to a terrestrial composition within about 5 Myr (Fig. 2b), so the isotopic similarity between Earth and the Moon can be explained if the giant impact involved bodies or precursors that completed their accretions towards the end of the disk's lifetime. The nearly identical isotope signatures of Earth and the Moon are thus an outcome of the evolving isotopic composition of the disk.

Online Content Methods, along with any additional Extended Data display items and Source Data, are available in the online version of the paper; references unique to these sections appear only in the online paper.

Received 28 July 2017; accepted 22 January 2018.

- Zhang, J., Dauphas, N., Davis, A. M., Leya, I. & Fedkin, A. The proto-Earth as a significant source of lunar material. *Nat. Geosci.* **5**, 251–255 (2012).
- Trinquier, A. *et al.* Origin of nucleosynthetic isotope heterogeneity in the solar protoplanetary disk. *Science* **324**, 374–376 (2009).
- Schiller, M., Paton, C. & Bizzarro, M. Evidence for nucleosynthetic enrichment of the protosolar molecular cloud core by multiple supernova events. *Geochim. Cosmochim. Acta* **149**, 88–102 (2015).
- Dauphas, N. The isotopic nature of the Earth's accreting material through time. *Nature* **541**, 521–524 (2017).
- Young, E. D. *et al.* Oxygen isotopic evidence for vigorous mixing during the Moon-forming giant impact. *Science* **351**, 493–496 (2016).
- Burkhardt, C. *et al.* In search of the Earth-forming reservoir: mineralogical, chemical, and isotopic characterizations of the ungrouped achondrite NWA 5363/NWA 5400 and selected chondrites. *Meteorit. Planet. Sci.* **52**, 807–826 (2017).
- Bollard, J. *et al.* Early formation of planetary building blocks inferred from Pb isotopic ages of chondrules. *Sci. Adv.* **3**, e1700407 (2017).
- Ciesla, F. J. Radial transport in the solar nebula: implications for moderately volatile element depletions in chondritic meteorites. *Meteorit. Planet. Sci.* **43**, 639–655 (2008).

- Lambrechts, M. & Johansen, A. Rapid growth of gas-giant cores by pebble accretion. *Astron. Astrophys.* **544**, A32 (2012).
- Johansen, A., Mac Low, M. M., Lacerda, P. & Bizzarro, M. Growth of asteroids, planetary embryos, and Kuiper belt objects by chondrule accretion. *Sci. Adv.* **1**, e1500109 (2015).
- Dauphas, N. *et al.* Calcium-48 isotopic anomalies in bulk chondrites and achondrites: evidence for a uniform isotopic reservoir in the inner protoplanetary disk. *Earth Planet. Sci. Lett.* **407**, 96–108 (2014).
- Currie, T. & Sicilia-Aguilar, A. The transitional protoplanetary disk frequency as a function of age: disk evolution in the Coronet Cluster, Taurus, and other 1–8 Myr old regions. *Astrophys. J.* **732**, 24 (2011).
- McSween, H. Y. *et al.* Dawn; the Vesta–HED connection; and the geologic context for eucrites, diogenites, and howardites. *Meteorit. Planet. Sci.* **48**, 2090–2104 (2013).
- Wang, H. *et al.* Lifetime of the solar nebula constrained by meteorite paleomagnetism. *Science* **355**, 623–627 (2017).
- Schiller, M., Baker, J. A. & Bizzarro, M. ^{26}Al – ^{26}Mg dating of asteroidal magmatism in the young Solar System. *Geochim. Cosmochim. Acta* **74**, 4844–4864 (2010).
- Keil, K. Angrites, a small but diverse suite of ancient, silica-undersaturated volcanic-plutonic mafic meteorites, and the history of their parent asteroid. *Chem. Erde* **72**, 191–218 (2012).
- Macke, R. J., Britt, D. T. & Consolmagno, G. J. Density, porosity, and magnetic susceptibility of achondritic meteorites. *Meteorit. Planet. Sci.* **46**, 311–326 (2011).
- Wilson, L., Goodrich, C. A. & Van Orman, J. A. Thermal evolution and physics of melt extraction on the ureilite parent body. *Geochim. Cosmochim. Acta* **72**, 6154–6176 (2008).
- Wilson, L. & Keil, K. Volcanic activity on differentiated asteroids: a review and analysis. *Chem. Erde* **72**, 289–321 (2012).
- Schiller, M. *et al.* Rapid timescales for magma ocean crystallization on the howardite–eucrite–diogenite parent body. *Astrophys. J. Lett.* **740**, L22 (2011).
- Schiller, M., Connolly, J. N., Glad, A. C., Mikouchi, T. & Bizzarro, M. Early accretion of protoplanets inferred from a reduced inner solar system ^{26}Al inventory. *Earth Planet. Sci. Lett.* **420**, 45–54 (2015).
- Bollard, J., Connolly, J. N. & Bizzarro, M. Pb–Pb dating of individual chondrules from the CBa chondrite Gujba: assessment of the impact plume formation model. *Meteorit. Planet. Sci.* **50**, 1197–1216 (2015).
- Dauphas, N. & Pourmand, A. Hf–W–Th evidence for rapid growth of Mars and its status as a planetary embryo. *Nature* **473**, 489–492 (2011).
- Hartmann, L., Herczeg, G. & Calvet, N. Accretion onto pre-main-sequence stars. *Annu. Rev. Astron. Astrophys.* **54**, 135–180 (2016).
- Kruijer, T. S., Burkhardt, C., Budde, G. & Kleine, T. Age of Jupiter inferred from the distinct genetics and formation times of meteorites. *Proc. Natl Acad. Sci. USA* **114**, 6712–6716 (2017).
- Picogna, G. & Kley, W. How do giant planetary cores shape the dust disk? HL Tauri system. *Astron. Astrophys.* **584**, A110 (2015).
- Carrasco-González, C. *et al.* The VLA view of the HL Tau disk: disk mass, grain evolution, and early planet formation. *Astrophys. J. Lett.* **821**, L16 (2016).
- Hartmann, W. K. & Davis, D. R. Satellite-sized planetesimals and lunar origin. *Icarus* **24**, 504–515 (1975).
- Canup, R. M. Lunar-forming impacts: processes and alternatives. *Phil. Trans. R. Soc. A* **372**, 20130175 (2014).
- Canup, R. M. Forming a Moon with an Earth-like composition via a giant impact. *Science* **338**, 1052–1055 (2012).
- Čuk, M. & Stewart, S. T. Making the Moon from a fast-spinning Earth: a giant impact followed by resonant despinning. *Science* **338**, 1047–1052 (2012).
- Meier, M. M. M., Reufer, A. & Wieler, R. On the origin and composition of Theia: constraints from new models of the Giant Impact. *Icarus* **242**, 316–328 (2014).
- Kleine, T. *et al.* Hf–W chronology of the accretion and early evolution of asteroids and terrestrial planets. *Geochim. Cosmochim. Acta* **73**, 5150–5188 (2009).

Acknowledgements Financial support for this project was provided to M.B. by the Danish National Research Foundation (DNRF97) and the European Research Council (ERC Consolidator Grant Agreement 616027—STARDUST2ASTEROIDS). V.A.F. acknowledges financial support from a DFG-Eigenstelle FE 1523/3-1 and the Royal Society for the purchase of Dhofar 287. We thank Å. Nordlund, A. Johansen and F. Moynier for discussion on the paper, as well as J. Day for comments that helped improve the quality of our paper.

Author Contributions M.S. and M.B. designed the study and experiments. M.S. conducted the analytical work. All authors participated in the interpretation of the data. M.S. and M.B. wrote the manuscript with input from V.A.F.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Publisher's note: Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations. Correspondence and requests for materials should be addressed to M.S. (schiller@snm.ku.dk).

Reviewer Information Nature thanks J. Day and the other anonymous reviewer(s) for their contribution to the peer review of this work.

METHODS

Sample preparation and isotope analyses. Bulk samples weighing 20–100 mg were digested in a HF-HNO₃ medium in Parr bombs at 210 °C for 2–3 days. Calcium was also separated from sample digestions of nine chondrules extracted from the ordinary chondrite (L3.10) NWA 5697 and two chondrules from the CR2 sample NWA 6043, which have previously been analysed for Pb isotopes⁷. The individual sample dissolutions typically represent less than 10 mg of sampled chondrule material, which is significantly smaller than the amount of processed bulk meteorite samples. Following complete dissolution of the samples, Ca was separated from the sample matrix by ion-exchange chromatography³⁴ in a four-step procedure. Although this technique is usually efficient, all samples were monitored for the presence of contaminant elements, such as Mg, Sr and Ti, in the purified-Ca solution. Individual chromatography separation steps were repeated when the presence of contaminants could affect the analysis (for example, when the Sr/Ca concentration ratio was higher than 1×10^{-6}). The isotopic compositions of the purified-Ca separates were measured with a Neptune Plus multiple-collector inductively coupled plasma mass spectrometer (MC-ICPMS) at the Centre for Star and Planet Formation (Natural History Museum of Denmark, University of Copenhagen) by following established analytical procedures^{3,34}. Data were acquired in the static mode using six Faraday collectors, one for each of the isotopes ⁴⁸Ca, ⁴⁷Ti, ⁴⁶Ca, ⁴²Ca, ⁴³Ca and ⁴⁴Ca. The Faraday cup used for ⁴⁷Ti was connected to an amplifier with a $10^{12} \Omega$ feedback resistor, whereas the other collectors were connected to amplifiers with $10^{11} \Omega$ feedback resistors. Samples were aspirated into the plasma source by means of an Apex sample introduction system with an uptake rate of $20 \mu\text{l min}^{-1}$, and the Ca isotopes were measured with a mass resolving power ($M/\Delta M$, as defined by the peak edge width from 5% to 95% at full peak height) that was always greater than 5,000. The sensitivity under these analytical conditions was approximately $300 \text{ V (p.p.m.)}^{-1}$.

Given the low abundance of the ⁴⁶Ca nuclide and the limited $\mu^{43}\text{Ca}$ variability (lower than 20 p.p.m.) between bulk inner-Solar-System reservoirs relative to the measurement uncertainty (about 2 p.p.m.)^{3,34}, we focus on the high-precision determination of the mass-independent ⁴⁸Ca/⁴⁴Ca ratio, $\mu^{48}\text{Ca}$, which we measure with an external reproducibility of 12 p.p.m. for individual sample analyses³⁴. We also report the mass-dependent ⁴²Ca/⁴⁴Ca ($\delta^{42/44}\text{Ca}$) and ⁴³Ca/⁴⁴Ca ($\delta^{43/44}\text{Ca}$) values with respect to SRM 915b, which are calculated in the same way as $\mu^{48}\text{Ca}$, but are not corrected for mass fractionation. For each sample of sufficient size, the analysis consists of 15–20 sample measurements, each of which comprises 839 s of data acquisition and 340 s of baseline measurements and is bracketed by measurements of the SRM 915b standard. For individual chondrules, both the signal intensity and the number of repeat analyses were adjusted to achieve the best possible precision, given the limited amount of Ca available in these samples. The data reduction was conducted off-line and changes in mass bias with time were interpolated using a smoothed cubic spline. For each analysis, the mean and standard error of the measured ratios were calculated using a three-standard-deviation threshold to reject outliers. Individual analyses of a sample were combined to produce an average weighted by the propagated uncertainties of individual analyses. The final uncertainties reported here are twice the standard error of the mean.

Calculation of disk-mass accretion rates. For simplicity, we assume that the mass of asteroids, moons and planets located sunwards of Jupiter represents the final mass of the protoplanetary disk and the $\mu^{48}\text{Ca}$ signature of Earth, which makes up 50% of this mass, is representative of the final inner-disk composition. This assumption is justified because $\mu^{48}\text{Ca}$ reveals the composition of disk solids, but not that of gas. The mass accreted in bodies at any given time is assumed to be traced by the bulk $\mu^{48}\text{Ca}$ signature of planetary bodies: increase in the $\mu^{48}\text{Ca}$ value of planetary bodies represents an addition of material from the outer Solar System to the growing inner disk. In this simple two-component mixing model, the ureilite parent body represents the mean composition of the inner disk 0.1 Myr after the formation of the Solar System. The $\mu^{48}\text{Ca}$ signature of CI chondrites represents the added outer-Solar-System material during the lifetime of the disk, and the $\mu^{48}\text{Ca}$ signature of Earth corresponds to the final composition of the disk after accretion has ceased. This mixing model requires that 58.4% of the material had accreted by the time the ureilite parent body formed (less than 0.1 Myr after the formation of the Solar System). The subsequent increase in the bulk $\mu^{48}\text{Ca}$ signatures of Vesta and the angrite parent body, Mars at half of its current mass and the proto-Earth represent addition of 15.0%, 19.6% and 6.9% outer-Solar-System material during the time intervals between the accretion of: 1) the ureilite parent body and Vesta (0.1–0.4 Myr), 2) Vesta and Mars at 50% of its present mass (0.4–2.7 Myr) and 3) Mars at 50% of its present mass and the proto-Earth (2.7–5.3 Myr), respectively. The amount of early-accreted material only increases if we assume the ⁴⁸Ca signature of carbonaceous chondrites other than CI chondrites. The change in accreted mass per time interval relative to the final disk mass (that is, 0.584, 0.150, 0.196

and 0.069 of the final inner-disk mass) allows us to calculate mean inner-disk mass accretion rates for each of the four intervals by dividing the added mass per time interval by its duration (0.1 Myr, 0.3 Myr, 2.3 Myr and 2.6 Myr, respectively). The resulting accretion rates are $5.8 \times 10^{-6} M_{\text{inner disk}} \text{ yr}^{-1}$, $5.0 \times 10^{-7} M_{\text{inner disk}} \text{ yr}^{-1}$, $8.5 \times 10^{-8} M_{\text{inner disk}} \text{ yr}^{-1}$ and $2.6 \times 10^{-8} M_{\text{inner disk}} \text{ yr}^{-1}$, respectively.

Timing of accretion of the ureilite parent body. The energy release from the decay of the short-lived ²⁶Al radionuclide (half-life, $t_{1/2} = 0.73 \text{ Myr}$) is the most important heat source driving differentiation on small planetary bodies that were accreted during the first few millions of years from the formation of the Solar System. Therefore, the timing of accretion of differentiated bodies can be determined by estimating the time needed for the parent body to achieve global melting from the energy released by ²⁶Al. This requires knowledge of the timing of differentiation, as well as the initial abundance of ²⁶Al in the body's precursor material. Recent studies have shown that canonical levels of ²⁶Al found in calcium–aluminium-rich inclusions (CAIs) are not representative of the majority of the matter that made up the terrestrial-planet-forming region^{21,35}. Instead, the ²⁶Al/²⁷Al abundance ratio in this region at the time of CAI formation probably had values between 1 and 1.6×10^{-5} (refs 21, 35). On the basis of ²⁶Al–²⁶Mg and ⁵³Mn–⁵³Cr systematics, initial magmatic differentiation of the ureilite parent body took place at $4,567.1 \pm 1.1 \text{ Myr ago}$ ^{36,37}, which is the same age as the age of our Solar System based on CAI data³⁸. Assuming that the initial ²⁶Al/²⁷Al abundance in the material from which the ureilite parent body formed was 1.3×10^{-5} and the precursor material was ordinary chondrite-like in terms of density and composition, thermal modelling requires that the accretion was completed by 0.1 Myr after CAI formation. This is constrained by assuming an ambient temperature of 550 K in the accretion region of the ureilite parent body at about 0.1 Myr (ref. 39) and requires the parent body to achieve an internal temperature of 1,553 K, equivalent to 30% partial melting¹⁸, within 1.1 Myr after CAI formation, which is the time allowed by the age uncertainty for magmatic differentiation^{36,37}. We note that this accretion age is consistent with independent estimates of the accretion time of the ureilite parent body⁴⁰.

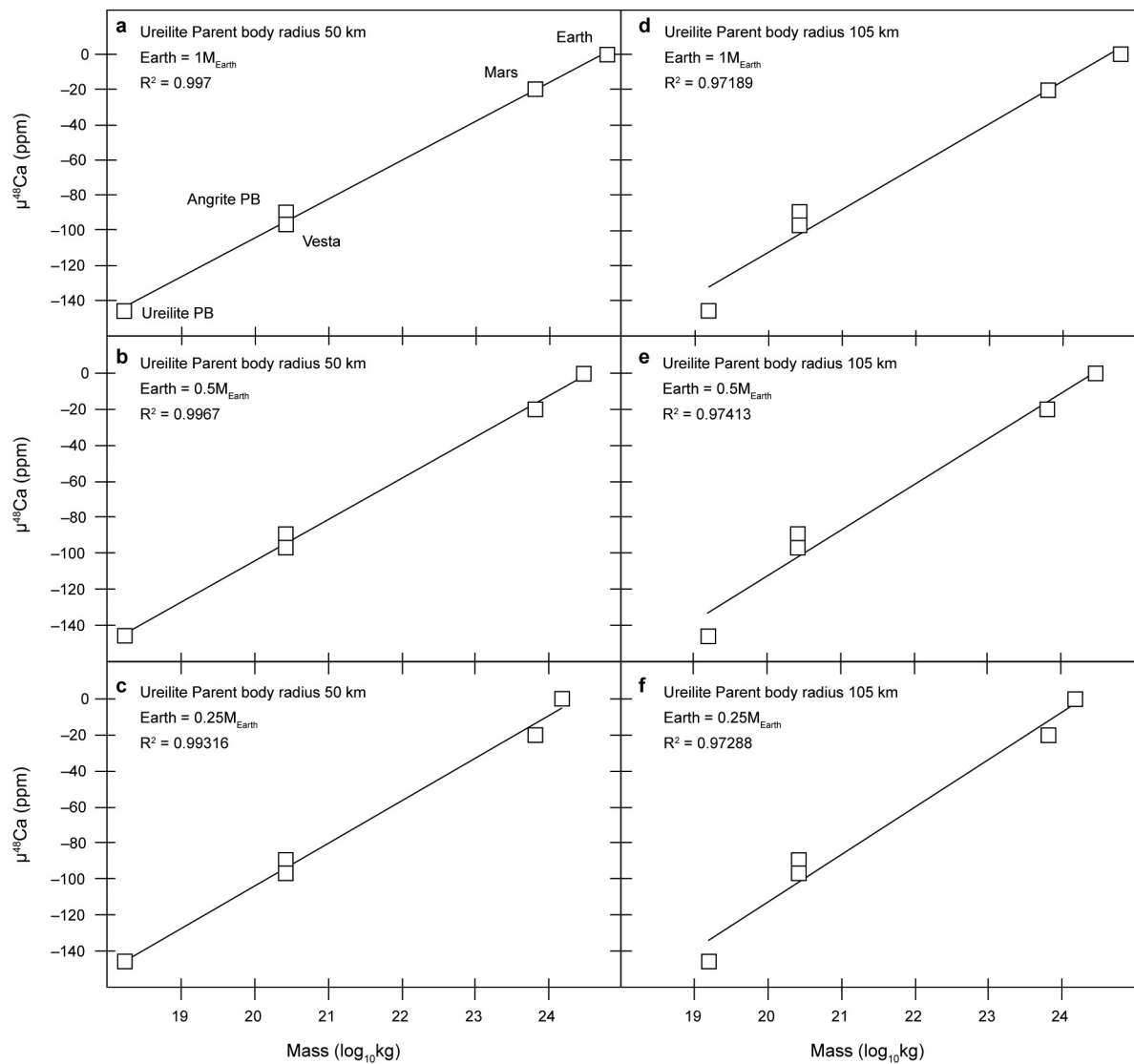
Stable Ca isotope composition of meteorites and chondrules. We also report the $\delta^{42/44}\text{Ca}$ and $\delta^{43/44}\text{Ca}$ values for our samples in Extended Data Table 1 and Extended Data Fig. 2 because mass-dependent fractionation can affect the accuracy of mass-independent $\mu^{48}\text{Ca}$ data via inappropriate corrections for natural mass fractionation. For example, correcting the natural mass-dependent fractionation by using a kinetic law to retrieve the $\mu^{48}\text{Ca}$ data will introduce an excess or deficit in $\mu^{48}\text{Ca}$ if the natural fractionation is driven by non-kinetic processes³⁴. Deviation from kinetic mass-dependent fractionation can occur during calcite precipitation⁴¹ and during chemical weathering in arid environments⁴²—the latter is relevant here because a number of our samples are hot-desert finds. Such corrections can change $\mu^{48}\text{Ca}$ by up to 12 p.p.m. for a 0.1‰ difference in the $\delta^{42/44}\text{Ca}$ value. The $\delta^{42/44}\text{Ca}$ values that we report here are in excellent agreement with literature data for the same group of meteorites (Extended Data Fig. 3)^{43–45}, confirming that the potential effects from such inadequate corrections are well within the uncertainty of our $\mu^{48}\text{Ca}$ results. Our $\delta^{42/44}\text{Ca}$ data for individual chondrules from ordinary and CR chondrites are also comparable to the value reported for CV chondrites in ref. 44. Moreover, we find no systematic differences (within uncertainty) in $\mu^{48}\text{Ca}$ between hot-desert and non-hot-desert meteorites, or between falls versus finds, in the same differentiated group (that is, ureilites, angrites and Mars) (Extended Data Fig. 4). Potential effects from the addition of terrestrial Ca into lunar and martian meteorites can also be ruled out on the basis of mass balance considerations. For example, assuming that the pristine martian $\mu^{48}\text{Ca}$ value is $-20 \pm 2.8 \text{ p.p.m.}$ relative to that of Earth, the addition of 15% of terrestrial Ca, which would be clearly identifiable from hand specimens, would result in a change of only 3 p.p.m. This effect is well within the external reproducibility of our measurements. Given that the ⁴⁸Ca composition of the Moon is even closer to that of Earth's, addition of 15% terrestrial Ca to a lunar meteorite would result in even smaller effects than those of martian meteorites. These data demonstrate that stable or mass-independent Ca isotope data are not affected by terrestrial processes before meteorite recovery, within the uncertainty of our measurements.

Other nucleosynthetic isotope data interpreted with our model. Nucleosynthetic anomalies have been reported for a number of elements besides Ca, such as Ti, Cr, Ni, Sr, Nd, Mo, Ru and O^{1,2,6,46–59}. Here we evaluate the consistency of our model with the isotope variability of these elements in various Solar System reservoirs. In this assessment, we consider whether it is possible to generate this nucleosynthetic variability in inner-Solar-System bodies by mixing a CI-like composition with a depleted inner-Solar-System dust with the composition of ureilite (where available) or angrite meteorites (Extended Data Fig. 5). On the basis of previous data^{1,2,6,46–54}, our model is compatible with the variability of isotopes of lithophile elements, such as ⁵⁰Ti, ⁵⁴Cr, ⁶²Ni and ¹⁴⁵Nd, within the uncertainties of the data (Extended Data Fig. 5). Nucleosynthetic anomalous isotopes (isotopes with nucleosynthetic

signatures that differ from Earth's) of other elements either do not show significant variability beyond typical analytical uncertainties (such as ^{84}Sr)^{55–57}, or do not track the entire accretion history of planetary bodies (siderophile elements, such as Mo and Ru)⁴, or may be affected by other processes, such as gas–water interactions (for example, O). Therefore, we conclude that our model is consistent with the nucleosynthetic variability of elements that track the source of the silicate fraction of asteroidal and planetary bodies.

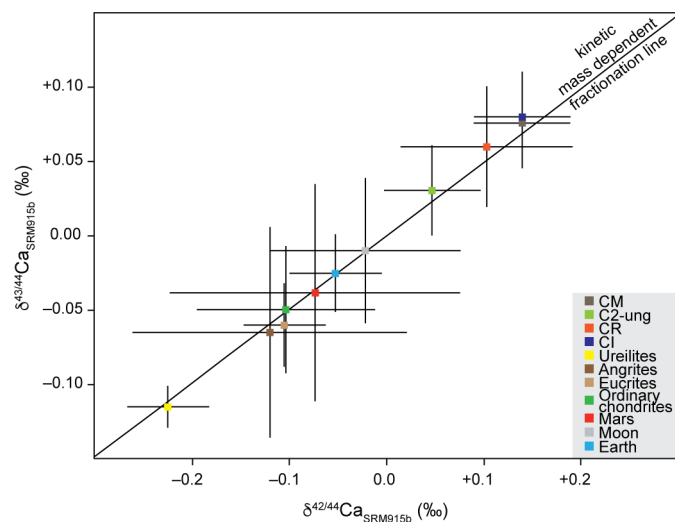
Data availability. Data supporting the findings of this study are available within the paper (including Methods and Extended Data) and from the EarthChem library (<http://dx.doi.org/10.1594/IEDA/100744>).

34. Schiller, M., Paton, C. & Bizzarro, M. Calcium isotope measurement by combined HR-MC-ICPMS and TIMS. *J. Anal. At. Spectrom.* **27**, 38–49 (2012).
35. Larsen, K. K. *et al.* Evidence for magnesium isotope heterogeneity in the solar protoplanetary disk. *Astrophys. J. Lett.* **735**, L37 (2011).
36. van Kooten, E. M., Schiller, M. & Bizzarro, M. Magnesium and chromium isotope evidence for initial melting by radioactive decay of ^{26}Al and late stage impact-melting of the ureilite parent body. *Geochim. Cosmochim. Acta* **208**, 1–23 (2017).
37. Baker, J. A., Schiller, M. & Bizzarro, M. ^{26}Al – ^{26}Mg deficit dating ultramafic meteorites and silicate planetesimal differentiation in the early Solar System? *Geochim. Cosmochim. Acta* **77**, 415–431 (2012).
38. Connelly, J. N. *et al.* The absolute chronology and thermal processing of solids in the solar protoplanetary disk. *Science* **338**, 651–655 (2012).
39. Estrada, P. R., Cuzzi, J. N. & Morgan, D. A. Global modeling of nebulae with particle growth, drift, and evaporation fronts. I. Methodology and typical results. *Astrophys. J.* **818**, 200 (2016).
40. Sugiura, N. & Fujiya, W. Correlated accretion ages and $\epsilon^{54}\text{Cr}$ of meteorite parent bodies and the evolution of the solar nebula. *Meteorit. Planet. Sci.* **49**, 772–787 (2014).
41. Gussone, N. *et al.* Calcium isotope fractionation in calcite and aragonite. *Geochim. Cosmochim. Acta* **69**, 4485–4494 (2005).
42. Ewing, S. A. *et al.* Non-biological fractionation of stable Ca isotopes in soils of the Atacama Desert, Chile. *Geochim. Cosmochim. Acta* **72**, 1096–1110 (2008).
43. Valdes, M. C., Moreira, M., Foriel, J. & Moynier, F. The nature of Earth's building blocks as revealed by calcium isotopes. *Earth Planet. Sci. Lett.* **394**, 135–145 (2014).
44. Amsellem, E. *et al.* Testing the chondrule-rich accretion model for planetary embryos using calcium isotopes. *Earth Planet. Sci. Lett.* **469**, 75–83 (2017); corrigendum 474 527 (2017).
45. Magna, T., Gussone, N. & Mezger, K. The calcium isotope systematics of Mars. *Earth Planet. Sci. Lett.* **430**, 86–94 (2015).
46. Tang, H. & Dauphas, N. Abundance, distribution, and origin of ^{60}Fe in the solar protoplanetary disk. *Earth Planet. Sci. Lett.* **359–360**, 248–263 (2012).
47. Tang, H. & Dauphas, N. ^{60}Fe – ^{60}Ni chronology of core formation in Mars. *Earth Planet. Sci. Lett.* **390**, 264–274 (2014).
48. Zhang, J., Dauphas, N., Davis, A. M. & Pourmand, A. A new method for MC-ICPMS measurement of titanium isotopic composition: identification of correlated isotope anomalies in meteorites. *J. Anal. At. Spectrom.* **26**, 2197–2205 (2011).
49. Trinquier, A., Birck, J. L. & Allegre, C. J. Widespread ^{54}Cr heterogeneity in the inner solar system. *Astrophys. J.* **655**, 1179–1185 (2007).
50. Yamakawa, A., Yamashita, K., Makishima, A. & Nakamura, E. Chromium isotope systematics of achondrites: chronology and isotopic heterogeneity of the inner solar system bodies. *Astrophys. J.* **720**, 150–154 (2010).
51. Qin, L., Alexander, C. M. D., Carlson, R. W., Horan, M. F. & Yokoyama, T. Contributors to chromium isotope variation of meteorites. *Geochim. Cosmochim. Acta* **74**, 1122–1145 (2010).
52. Carlson, R. W., Boyet, M. & Horan, M. Chondrite barium, neodymium, and samarium isotopic heterogeneity and early earth differentiation. *Science* **316**, 1175–1178 (2007).
53. Boyet, M. & Carlson, R. W. ^{142}Nd evidence for early (> 4.53 Ga) global differentiation of the silicate Earth. *Science* **309**, 576–581 (2005).
54. Borg, L. E., Brennecka, G. A. & Symes, S. J. Accretion timescale and impact history of Mars deduced from the isotopic systematics of martian meteorites. *Geochim. Cosmochim. Acta* **175**, 150–167 (2016).
55. Moynier, F. *et al.* Planetary-scale strontium isotopic heterogeneity and the age of volatile depletion of early Solar System materials. *Astrophys. J.* **758**, 45 (2012).
56. Paton, C., Schiller, M. & Bizzarro, M. Identification of an ^{84}Sr -depleted carrier in primitive meteorites and implications for thermal processing in the solar protoplanetary disk. *Astrophys. J.* **763**, L40 (2013).
57. Hans, U., Kleine, T. & Bourdon, B. Rb–Sr chronology of volatile depletion in differentiated protoplanets: BABI, ADOR and ALL revisited. *Earth Planet. Sci. Lett.* **374**, 204–214 (2013).
58. Clayton, N. R. Oxygen isotopes in meteorites. *Annu. Rev. Earth Planet. Sci.* **21**, 115–149 (1993).
59. Fischer-Gödde, M., Burkhardt, C., Kruijer, T. S. & Kleine, T. Ru isotope heterogeneity in the solar protoplanetary disk. *Geochim. Cosmochim. Acta* **168**, 151–171 (2015).

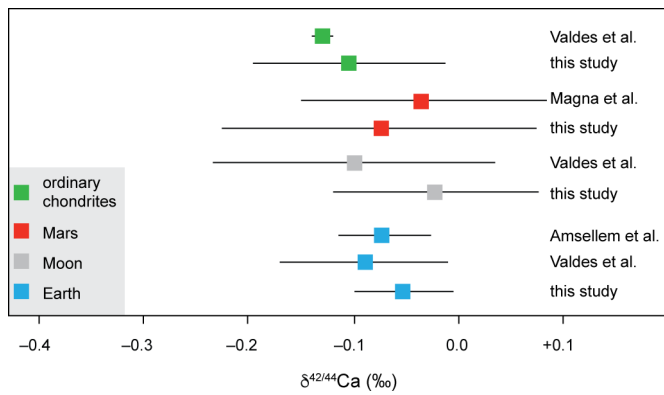


Extended Data Figure 1 | $\delta^{48}\text{Ca}$ values of planetary bodies versus mass for different masses of the ureilite parent body and Earth's precursor. a–c, Regressions (solid lines) and associated correlation coefficients through the data (squares) by assuming an ureilite parent body with a radius of 50 km and masses of M_{Earth} , $0.5M_{\text{Earth}}$ and $0.25M_{\text{Earth}}$ for Earth's

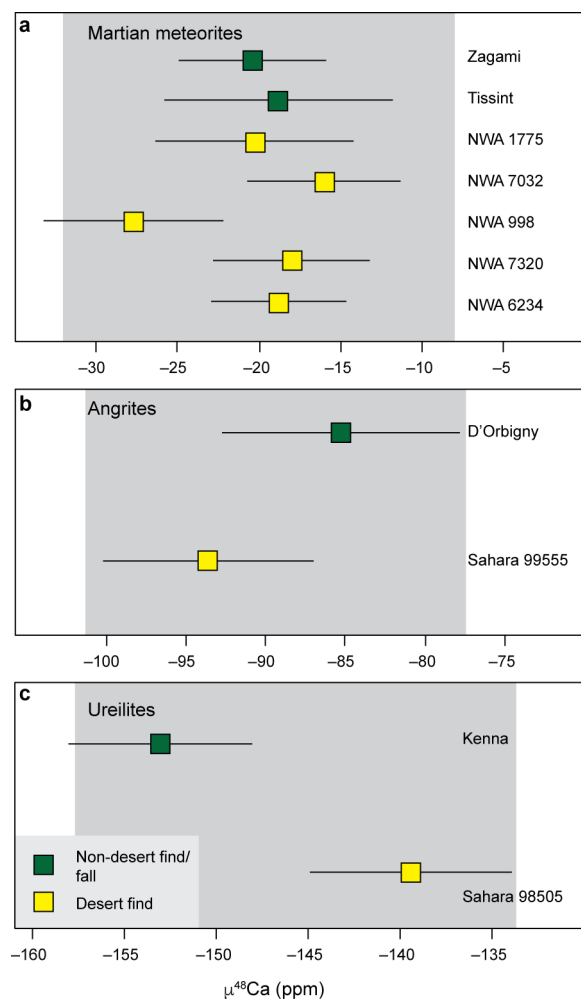
precursor. d–f, Regressions through the data but assuming a ureilite parent body with a radius of 105 km and masses of M_{Earth} , $0.5M_{\text{Earth}}$ and $0.25M_{\text{Earth}}$ for Earth's precursor. The masses for the angrite parent body, Vesta and Mars are the same as in Fig. 2a.



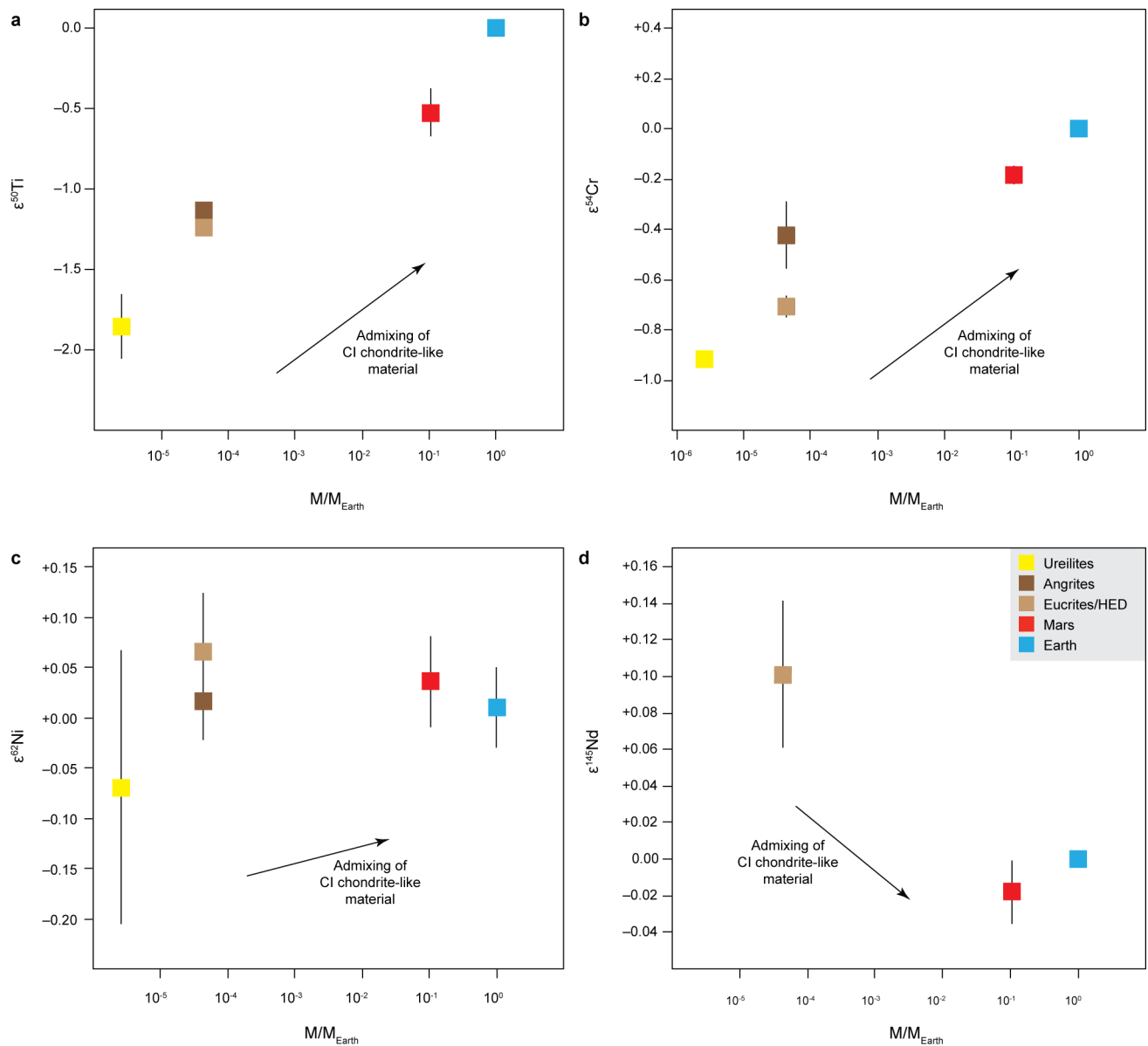
Extended Data Figure 2 | Three-isotope plot of the average $\delta^{42/44}\text{Ca}$ versus $\delta^{43/44}\text{Ca}$ for Earth, meteorite parent bodies and chondrite groups relative to the standard SRM 915b. The solid line shows the mass-dependent fractionation predicted by kinetic mass fractionation. Uncertainties shown for $\delta^{42/44}\text{Ca}$ and $\delta^{43/44}\text{Ca}$ are two times the standard error of the mean per group of analysed samples. For groups containing a single sample (ordinary chondrites, CI, CM and C2-ung), the error represents either the external reproducibility (0.05 and 0.03 for $\delta^{42/44}\text{Ca}$ and $\delta^{43/44}\text{Ca}$, respectively) or the analytical uncertainty of the measurement; whichever is larger.



Extended Data Figure 3 | Comparison of $\delta^{42/44}\text{Ca}$ values with previous results. Data for ordinary chondrites and martian, lunar and terrestrial basalts^{43–45} are compared with those determined in this study. Uncertainties shown are two standard errors of the mean.



Extended Data Figure 4 | Comparison of $\mu^{48}\text{Ca}$ values determined for desert and non-desert finds or falls. Data are shown for martian (a), angrite (b) and ureilite (c) meteorites. The grey shaded area indicates the external reproducibility of individual sample analyses. Uncertainties shown are two standard errors of the mean.



Extended Data Figure 5 | Correlation between parent-body mass and nucleosynthetic anomalies for ^{50}Ti , ^{54}Cr , ^{62}Ni and ^{145}Nd . The data are from refs 1, 2, 6, 46–54. The masses are shown relative to the mass of Earth, M_{Earth} . Arrows indicate the effects of mixing CI-like matter with

the inner-disk reservoir on the isotope composition, as predicted on the basis of measured nucleosynthetic signatures of CI chondrites. Error bars indicate the 95% confidence level of the mean.

Extended Data Table 1 | Mass-independent $\mu^{48}\text{Ca}$ values and mass-dependent $\delta^{42/44}\text{Ca}$ and $\delta^{43/44}\text{Ca}$ data relative to SRM 915b

| Sample | $\mu^{48}\text{Ca} \pm 2\text{SE}$ | $\delta^{42/44}\text{Ca} \pm 2\text{SE}$ | $\delta^{43/44}\text{Ca} \pm 2\text{SE}$ |
|--------------------------------|------------------------------------|------------------------------------------|------------------------------------------|
| Earth | | | |
| BIR-1a (n=6) | +0.2 \pm 1.7 | −0.04 \pm 0.02 | −0.02 \pm 0.01 |
| BIR-1a† | +6.1 \pm 5.7 | −0.04 \pm 0.02 | −0.01 \pm 0.02 |
| BHVO-2† | +6.3 \pm 7.5 | −0.04 \pm 0.04 | −0.02 \pm 0.02 |
| BCR-2† | +5.2 \pm 3.4 | −0.08 \pm 0.04 | −0.04 \pm 0.02 |
| DNC-1† | +1.0 \pm 4.0 | −0.04 \pm 0.03 | −0.02 \pm 0.01 |
| BHVO-1† | −6.1 \pm 8.1 | −0.08 \pm 0.03 | −0.04 \pm 0.02 |
| SGR-1† | −4.6 \pm 8.7 | −0.02 \pm 0.02 | −0.01 \pm 0.01 |
| Sarm 40† | −6.7 \pm 9.7 | −0.08 \pm 0.02 | −0.04 \pm 0.01 |
| Mean | +0.2 \pm 3.9 | −0.05 \pm 0.02 | −0.03 \pm 0.01 |
| Moon | | | |
| NWA 4734 (n=7) | +3.4 \pm 6.7 | −0.08 \pm 0.04 | −0.04 \pm 0.02 |
| Dhofar 026 (n=2) | +6.4 \pm 2.5 | +0.03 \pm 0.00 | +0.02 \pm 0.00 |
| NWA 8632 (n=2) | +3.2 \pm 3.3 | +0.00 \pm 0.13 | +0.00 \pm 0.06 |
| Dhofar 287 | +1.9 \pm 4.7 | −0.04 \pm 0.03 | −0.02 \pm 0.02 |
| Mean | +3.7 \pm 1.9 | −0.02 \pm 0.05 | −0.01 \pm 0.02 |
| Mars | | | |
| NWA 6234 | −18.8 \pm 4.2 | −0.03 \pm 0.01 | −0.01 \pm 0.00 |
| NWA 7320 | −18.0 \pm 4.8 | −0.03 \pm 0.02 | −0.02 \pm 0.01 |
| NWA 998 | −27.7 \pm 5.5 | −0.01 \pm 0.01 | −0.01 \pm 0.00 |
| NWA 7032 | −16.0 \pm 4.7 | −0.02 \pm 0.01 | −0.02 \pm 0.01 |
| NWA 1775 | −20.3 \pm 6.1 | −0.08 \pm 0.04 | −0.04 \pm 0.02 |
| Tissint | −18.8 \pm 7.0 | −0.22 \pm 0.04 | −0.11 \pm 0.02 |
| Zagami | −20.4 \pm 4.5 | −0.12 \pm 0.04 | −0.06 \pm 0.02 |
| Mean | −20.0 \pm 2.8 | −0.07 \pm 0.06 | −0.04 \pm 0.03 |
| Ordinary chondrites | | | |
| Kramer Creek (L4) (n=4) | −34.5 \pm 10.7 | −0.14 \pm 0.07 | −0.07 \pm 0.03 |
| Bovedy (L3)* | −35.4 \pm 2.9 | −0.13 \pm 0.01 | −0.06 \pm 0.01 |
| NWA 4910 (LL3) | −44.3 \pm 9.3 | −0.04 \pm 0.01 | −0.02 \pm 0.01 |
| NWA 5697 (L3) (n=3) | −25.5 \pm 8.2 | −0.11 \pm 0.03 | −0.05 \pm 0.01 |
| Mean | −34.9 \pm 7.7 | −0.10 \pm 0.05 | −0.05 \pm 0.02 |
| Eucrites | | | |
| Juvinas* | −91.5 \pm 7.1 | −0.12 \pm 0.03 | −0.07 \pm 0.02 |
| Stannern* | −102.4 \pm 4.3 | −0.09 \pm 0.02 | −0.05 \pm 0.01 |
| Mean | −97.0 \pm 10.9 | −0.11 \pm 0.03 | −0.06 \pm 0.02 |
| Angrites | | | |
| D'Orbigny* | −85.3 \pm 7.5 | −0.07 \pm 0.02 | −0.04 \pm 0.01 |
| Sahara 99555* | −93.6 \pm 6.6 | −0.17 \pm 0.08 | −0.09 \pm 0.04 |
| Mean | −89.5 \pm 8.3 | −0.12 \pm 0.10 | −0.07 \pm 0.05 |
| Ureilites | | | |
| Kenna* | −153.0 \pm 5.0 | −0.21 \pm 0.01 | −0.11 \pm 0.01 |
| Sahara 98505* | −139.4 \pm 5.5 | −0.24 \pm 0.03 | −0.12 \pm 0.02 |
| Mean | −146.2 \pm 13.6 | −0.23 \pm 0.03 | −0.12 \pm 0.01 |
| Carbonaceous chondrites | | | |
| Ivuna (CI)* | +206.1 \pm 8.5 | +0.14 \pm 0.04 | +0.08 \pm 0.02 |
| EET 92161 (CR2) | +222.5 \pm 4.0 | +0.07 \pm 0.02 | +0.05 \pm 0.01 |
| NWA 1180 (CR2) | +206.6 \pm 4.6 | +0.13 \pm 0.01 | +0.07 \pm 0.00 |
| Tagish Lake (C2-ung) | +291.3 \pm 5.3 | +0.05 \pm 0.01 | +0.03 \pm 0.01 |
| Jbilet Winselwan (CM2) | +314 \pm 14 | +0.14 \pm 0.01 | +0.08 \pm 0.00 |
| Chondrules | | | |
| NWA 5697 (LL3) | | | |
| 2-C1 | −150 \pm 59 | −0.20 \pm 0.01 | −0.08 \pm 0.02 |
| 5-C2 | −145 \pm 53 | −0.31 \pm 0.10 | −0.11 \pm 0.04 |
| 5-C10 | −51 \pm 70 | −0.19 \pm 0.04 | −0.08 \pm 0.02 |
| D-C3 | −154 \pm 10 | +0.02 \pm 0.05 | +0.01 \pm 0.03 |
| 5-C4 | −15 \pm 31 | −0.33 \pm 0.04 | −0.15 \pm 0.02 |
| 3-C5 | −151 \pm 16 | −0.12 \pm 0.04 | −0.01 \pm 0.00 |
| 11-C1 | +16 \pm 33 | −0.19 \pm 0.03 | −0.08 \pm 0.02 |
| 11-C2 | −119 \pm 24 | −0.13 \pm 0.01 | −0.05 \pm 0.02 |
| 3-C2 | −101 \pm 19 | −0.10 \pm 0.01 | −0.05 \pm 0.01 |
| NWA 6043 (CR2) | | | |
| 1-C2 | +227 \pm 16 | −0.10 \pm 0.01 | −0.04 \pm 0.01 |
| 2-C4 | +198 \pm 24 | −0.06 \pm 0.01 | −0.04 \pm 0.01 |

$\mu^{48}\text{Ca}$ is expressed in parts per million, and $\delta^{42/44}\text{Ca}$ and $\delta^{43/44}\text{Ca}$ are in parts per thousand.
SE, standard error.

*Data from ref. 3.

†Data from ref. 34.

Humans thrived in South Africa through the Toba eruption about 74,000 years ago

Eugene I. Smith¹, Zenobia Jacobs², Racheal Johnsen¹, Minghua Ren¹, Erich C. Fisher^{3,4}, Simen Oestmo⁴, Jayne Wilkins⁵, Jacob A. Harris³, Panagiotis Karkanas⁶, Shelby Fitch¹, Amber Ciravolo¹, Deborah Keenan⁷, Naomi Cleghorn^{4,8}, Christine S. Lane⁹, Thalassa Matthews¹⁰ & Curtis W. Marean^{3,4}

Approximately 74 thousand years ago (ka), the Toba caldera erupted in Sumatra. Since the magnitude of this eruption was first established, its effects on climate, environment and humans have been debated¹. Here we describe the discovery of microscopic glass shards characteristic of the Youngest Toba Tuff—ashfall from the Toba eruption—in two archaeological sites on the south coast of South Africa, a region in which there is evidence for early human behavioural complexity. An independently derived dating model supports a date of approximately 74 ka for the sediments containing the Youngest Toba Tuff glass shards. By defining the input of shards at both sites, which are located nine kilometres apart, we are able to establish a close temporal correlation between them. Our high-resolution excavation and sampling technique enable exact comparisons between the input of Youngest Toba Tuff glass shards and the evidence for human occupation. Humans in this region thrived through the Toba event and the ensuing full glacial conditions, perhaps as a combined result of the uniquely rich resource base of the region and fully evolved modern human adaptation.

We use a date of approximately 74 ka as the time of eruption of the Youngest Toba Tuff (YTT), just before the onset of glacial marine isotope stage 4 (MIS4)^{2,3}. Early modern humans evolved in Africa before 100 ka and their main dispersal out of Africa began at about 70 ka⁴. To examine the potential effect of the Toba eruption on early modern humans⁵ requires the identification of the YTT signature directly intermixed with the behavioural traces of modern humans in Africa, which we have achieved in the sites discussed here.

Tephra—fragmented material from an explosive volcanic eruption—ranges in size from macroscopic layers to microscopic glass shards of ‘cryptotephra’ that are deposited in sediments⁶. We discovered YTT glass shards in archaeological sediments at the sites of Vleesbaai area B (VBB) and Pinnacle Point Site 5-6 (PP5-6) on the south coast of South Africa (Fig. 1 and Extended Data Fig. 1). To our knowledge, this is the first successful identification and analysis of shards at such a great distance (approximately 9,000 km) from a source volcano; the identification of shards at both sites, located 8.7 km apart, provides an unusually high-precision temporal link between an open-air site and a rock shelter that were probably used by the same group of people.

The published dates for the PP5-6 sequence range from approximately 90 to 50 ka, and are based on single-grain optically stimulated luminescence (OSL) dating⁷. We present an updated age model (Fig. 2) for PP5-6, constructed from 90 OSL-dated sediment samples analysed with Bayesian statistics, and calculated on the OxCal platform (<https://c14.arch.ox.ac.uk/oxcal.html>, see Methods and Supplementary

Information), for comparison to the distribution of YTT glass shards (see Methods). The rapid deposition and repeated occupation of the site by humans has created an approximately 15-vertical-metre, high-resolution record of human occupation across the timespan of the Toba eruption. The sediments at this site are grouped into large sedimentary units called ‘aggregates’; thinner layers within these aggregates are referred to as ‘sub-aggregates’⁷. The aggregates of interest here, from bottom to top, include ‘light brown sand and roofspall’ (LBSR), which is dated to about 90–74 ka, ‘aeolian light brown sand’ (ALBS) and ‘shelly ashy brown sand’ (SADBS), which together date to around 74–71 ka⁷ (Fig. 2). We used total stations (see Methods) to measure the 3D location of over 400,000 archaeological finds to millimetre accuracy⁸, which enables precise investigation of the relationship of the shards to the archaeological evidence for changes in the intensity of human occupation. We collected continuous sediment samples (2–3 g per sample) across these aggregates to search for shards (Fig. 2, Extended Data Figs 2–4 and Supplementary Videos 1–4).

At Vleesbaai, Later and Middle Stone Age occupations occur in stacked ancient dunes and palaeosols⁹. Excavations at VBB began in 2015, and have revealed lithic and faunal finds sealed below a post-depositional cemented layer. We excavated an area of 14 m² and recovered around 900 artefacts, which were plotted by total station and which came from a single lens that is limited to about a 10–15 cm vertical spread (Fig. 3). OSL and sediment samples were collected from an adjacent geological trench, located 6.8 m away from the archaeological excavation (Fig. 3). Two OSL samples from the artefact-bearing sediment, the aggregate ‘dark yellowish-brown clayey sand’ (DYBCS), give estimates dating to 73.4 ± 5.2 and 73.9 ± 4.2 ka. Two samples from the overlying ‘light yellowish brown sand 2’ (LYBS2) aggregate gave dates of 72.5 ± 4.3 and 71.4 ± 4.5 ka. The underlying ‘strong brown clayey sand’ (SBCS) aggregate was dated to 74.7 ± 5.8 ka, and the lowest unit—the ‘very pale brown sand and concretions’ (VPBSC) aggregate—was dated to 98.7 ± 7.6 ka. To determine whether cryptotephra shards were present, we collected continuous sediment samples at 1-cm intervals from a vertical 1.5-m section of the geological trench, across the stratigraphic contacts between DYBCS, SBCS and underlying VPBSC.

The identification of the source and age of cryptotephra is a tool used to define isochrons for correlation of sediments across sites and regions⁶. Samples were processed using cryptotephra separation techniques^{6,10} modified for ‘extremely low abundance’ cryptotephra (Supplementary Information). Major elements were analysed by electron microprobe, and trace elements were analysed by laser-ablation inductively coupled plasma mass spectrometry (LA-ICP-MS)

¹Department of Geoscience, University of Nevada Las Vegas, 4505 Maryland Parkway, Las Vegas, Nevada 89154, USA. ²ARC Centre of Excellence for Australian Biodiversity and Heritage & Centre for Archaeological Science, School of Earth and Environmental Sciences, University of Wollongong, Wollongong, New South Wales 2522, Australia. ³Institute of Human Origins, School of Human Evolution and Social Change, PO Box 872402, Arizona State University, Tempe, Arizona 85287-2402, USA. ⁴African Centre for Coastal Palaeoscience, Nelson Mandela University, Port Elizabeth, Eastern Cape 6031, South Africa. ⁵Human Evolution Research Institute, Department of Archaeology, University of Cape Town, Private Bag, Rondebosch 7701, South Africa. ⁶Malcolm H. Wiener Laboratory for Archaeological Science, American School of Classical Studies, Soudias 54, Athens 10676, Greece. ⁷Geoscience Consultants LLC, Henderson, Nevada 89014, USA. ⁸Department of Sociology and Anthropology, University of Texas at Arlington, 701 South Nedderman Drive, Arlington, Texas 76019, USA. ⁹Department of Geography, University of Cambridge, Downing Place, Cambridge, CB2 3EN, UK. ¹⁰Iziko Museums of South Africa, Queen Victoria Street, Cape Town, PO Box 61, Cape Town, 8000, South Africa.

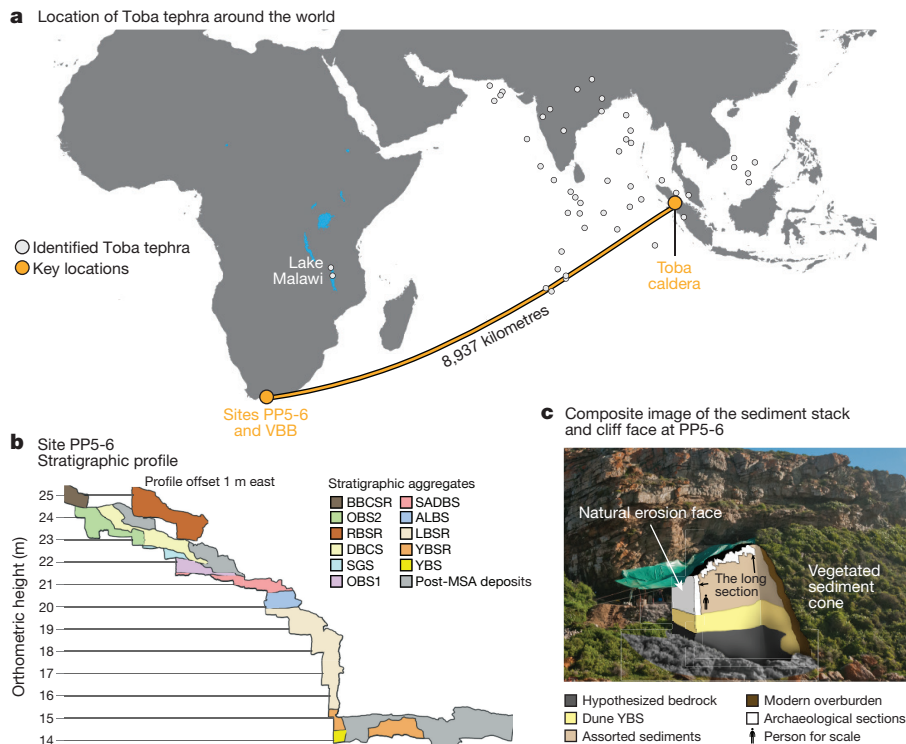


Figure 1 | VBB and PP5-6 and its relationship to other YTT study sites. **a**, Location of VBB and PP5-6 relative to the Toba caldera and other locations at which YTT ash has been found. **b**, Schematic profile of the stratigraphic sequence at PP5-6 that shows the major depositional changes. **c**, Composite view of PP5-6 that shows an interpretation of how

the sediment stack developed in front of the cliff face, and its relation to the 'long section' excavation area. BBCSR, brown and black compact sand and roofspall; DBCS, dark brown compact sand; OBS, orange brown sand; RBSR, reddish brown sand and roofspall; SGS, shelly grey sand; YBS, yellowish brown sand; YBSR, yellowish brown sand and roofspall.

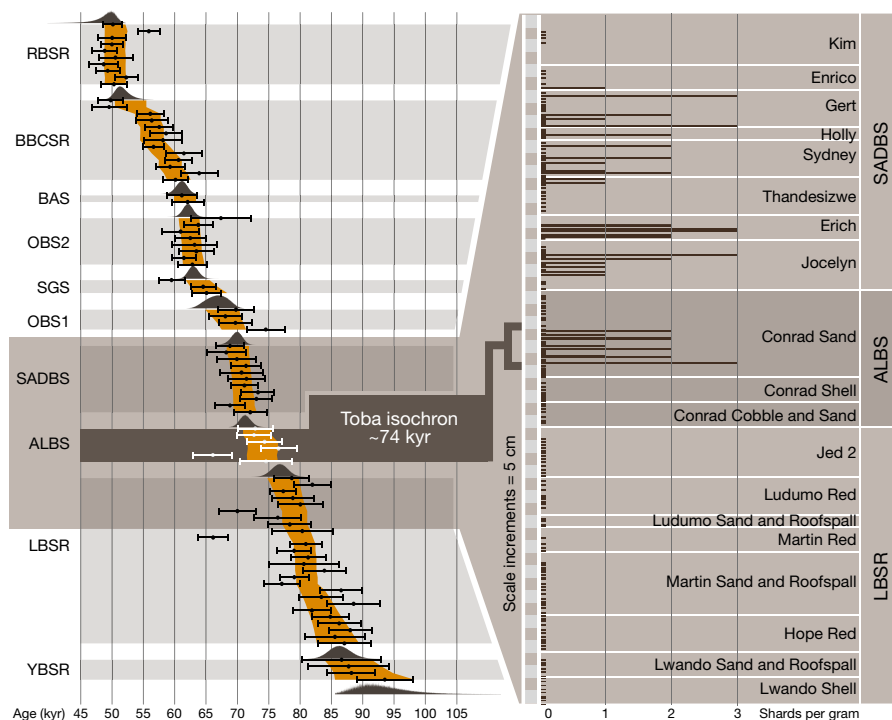


Figure 2 | The location of the YTT isochron at PP5-6. Left, current PP5-6 age model developed using OxCal v.4.2 (<https://c14.arch.ox.ac.uk/oxcal.html>) from OSL ages. Individual OSL ages are shown with their standard errors. The modelled boundaries between the stratigraphic aggregates are represented using the 95.4% confidence interval, whereas the continuous age model is represented at the 68% confidence interval. Right, details at the LBSR-ALBS and the ALBS-SADBS contacts, including

the distribution of YTT glass shards throughout this sequence for all shard sample transects. The shard counts are compiled by calculating the relative height of each sample above the lowermost stratigraphic sub-aggregate contact along each transect line, to compensate for lateral topographic variations across the sections. The YTT isochron is identified in the ALBS Conrad Sand. kyr, thousand years; BAS, black ashy sand.

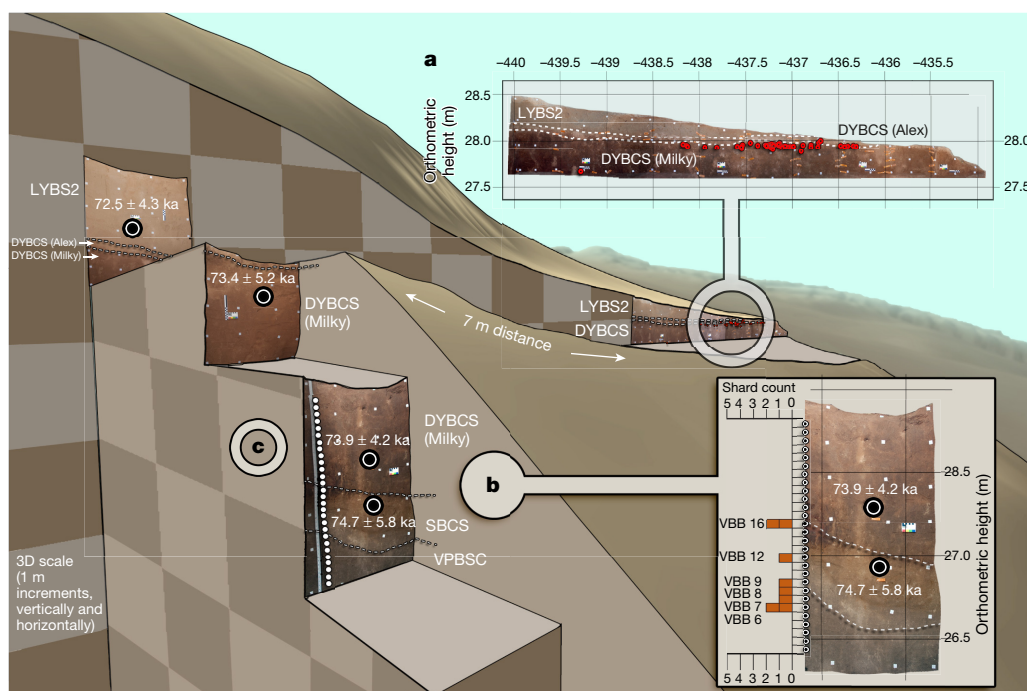


Figure 3 | The shard distribution, OSL dates and artefact plots as a composite digital cutaway at VBB. a, VBB site stratigraphy and 3D plotted artefacts on a digitally rectified stratigraphic photograph. b, Shard distribution on a digitally rectified stratigraphic photograph. c, Composite

digital cutaway of VBB dune deposits showing stacked stratigraphic profiles in the north-south geological trench, and the east-west stratigraphic profile in the nearby archaeological excavations. OSL dating samples are indicated with black-and-white dots.

or secondary ion mass spectrometry (SIMS) (see Methods). We searched the databases of the 'Volcano Global Risk Identification and Analysis Project' (VOGRIPA; <http://www.bgs.ac.uk/vogripa/>), the Smithsonian Institution's 'Global Volcanism Program' (<https://volcano.si.edu/>) and the Petrological Database (<http://www.earthchem.org/petdb>) for moderate-to-large rhyolitic eruptions that occurred between 90 and 40 ka, and used these as the bases of comparison for all available compositional data. Only the YTT produced tephra that matched the major- and rare-earth-element chemistry of cryptotephra at VBB and PP5-6, and was of the proper age to contribute tephra to these sites (Fig. 4, Extended Data Figs 5, 6, Extended Data Tables 1, 2 and Supplementary Tables 1-5).

The major-element chemistry of cryptotephra from VBB is nearly identical to that of distal YTT previously described from India¹¹ and Lake Malawi¹² and to trace-element abundances in distal YTT glass¹¹ and rhyolite from the Toba caldera¹³. Cryptotephra from PP5-6 are a close match for YTT, but are slightly offset (see Supplementary Discussion). The shards from PP5-6 are in sediments below the drip line at PP5-6. This zone is a complex depositional environment⁷, and these shards may therefore have been subjected to chemical alteration that has resulted in the chemical differences between PP5-6 cryptotephra and typical distal YTT cryptotephra.

We built a Bayesian model to formally assign the probability that individual shards belong to a known eruption (see Methods). When applied to shards from VBB and PP5-6, the model shows that, for each shard analysis, there is a >96% probability that they correlate with YTT (Extended Data Tables 3, 4, Methods and Supplementary Tables 8-10). Despite the slight chemical offset of the PP5-6 shards, the chemical plots and Bayesian model show that these shards more closely match YTT than any other known eruption. The fact that shards at VBB are a near-perfect match to YTT cryptotephra, and that OSL dates for the sediments at both sites correspond to the age of the Toba eruption, leads us to conclude that Toba is the source volcano for both sets of shards. We recognize that the chemistry of the shards at PP5-6 is not a perfect match, probably as a result of post-depositional alteration. To our knowledge, these shards are

the farthest-travelling shards from any volcano that have yet been discovered.

We extracted and counted shards from the sediment samples to create a shard profile. The cryptotephra at PP5-6 display a vertical shard profile with a few sharp distinct peaks that are concentrated in ALBS sediments above the LBSR, with a fluctuating but generally diminishing shard content as one moves up the stratigraphy (Fig. 2). This pattern is typical of the near-instantaneous deposition of material after a volcanic eruption⁶. The first appearance and highest concentration of shards occurs near the top of an aeolian sand unit, the sub-aggregate 'Conrad Sand'. This first peak is overlain by an approximately 10-cm-thick layer of aeolian sand that is free of shards; this layer is sealed by a series of successive intact hearth features in the lower portion of 'Jocelyn' sub-aggregate that are also devoid of shards (Fig. 2). We place the YTT isochron at the position of Conrad Sand (Fig. 2 and Extended Data Fig. 4).

The transition from roofspall-dominated (LBSR) to aeolian-dominated (ALBS) sediments (Extended Data Fig. 4) indicates a major environmental change that reflects the onset of MIS4 and retreat of the sea⁷. The ALBS aeolian unit has within it one dense human occupation layer ('Conrad Shell') and several thin, less-dense human occupation lenses. The first tephra peak is located just above the Conrad Shell layer, and is interstratified with aeolian sand and evidence of human occupation. Our Bayesian age model dates the transition between LBSR and ALBS to 79.5-74.1 ka (95.4% probability for all dates). The YTT isochron in the Conrad Sand has two measured OSL dates of 72.6 ± 5.6 and 72.8 ± 5.8 ka, with Bayesian-modelled time intervals of 75.4-70.9 ka, both of which are statistically consistent with the approximately 74 ka estimate for YTT^{2,3}.

At VBB, the first occurrence of the shards is at the top of VPBCS, near where this aggregate contacts SBCS. A second occurrence in the shard profile occurs at the contact between SBCS and DYBCS (Fig. 3). No shards were recovered from sediments above this contact or below the top of VPBCS (Fig. 3). The cryptotephra-bearing sediments at VBB have OSL dates of between 73.9 ± 4.2 and 74.7 ± 5.8 ka and thus are also statistically consistent with the age of YTT. The artefacts were

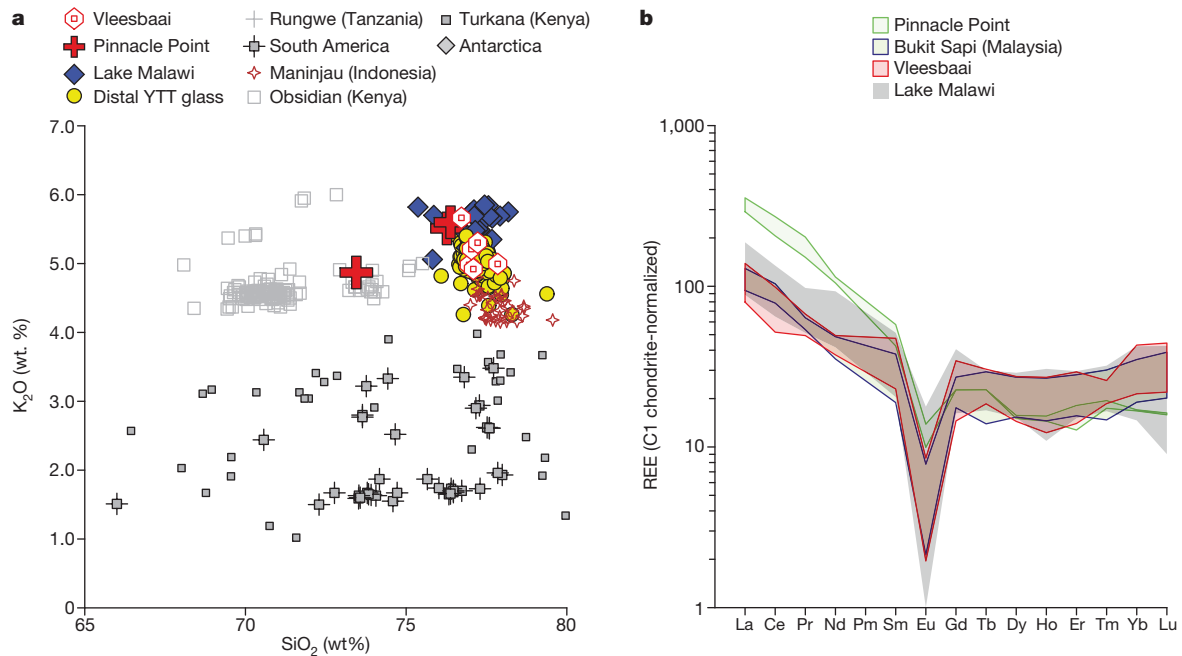


Figure 4 | Geochemical comparisons between extremely low abundance cryptotephra at VBB and PP5-6, and distal and proximal YTT.

a, b, K_2O versus SiO_2 (weight per cent, wt%) (a) and rare-earth elements (REE) normalized to chondrite (b) comparing VBB and PP5-6 shards to YTT glass. Gd and Tb were not analysed by SIMS. In **a**, tephra from

other volcanoes in Indonesia (Maninjau), Africa (Rungwe and Turkana), Antarctica and South America (see Supplementary Information) are also plotted. In **a**, samples from Rungwe and Antarctica have SiO_2 values of less than 65 wt% and are not shown on the plots.

recovered from the top of DYBCS, about 1.5 m above the highest shard in the vertical profile. The OSL dates of 73.9 ± 4.2 and 73.4 ± 5.2 ka for the bottom and top of DYBCS, respectively, are consistent with the deposition of the VBB artefacts at around 74 ka, at—or shortly after—the deposition of YTT cryptotephra.

Finding YTT in both the PP5-6 and VBB sedimentary sequences illustrates the transformative potential in being able to temporally link multiple sites, including open-air focused activity sites and home-base sites such as rock shelters or caves. The Toba event has previously been located in both the Greenland and Antarctic ice cores, where synchronous volcanic markers have been identified within Greenland interstadial 20 (Antarctic isotope maxima 19 and 20)¹⁴, and thus form an isochron from archaeological sites in South Africa to Greenland, around 13,000 km apart.

At PP5-6, the LBSR-to-ALBS transition and YTT isochron accompanies changes from low-intensity human occupation (in LBSR) to high-intensity occupation coupled with intense anthropogenic burning (in SADBS)⁷. The frequency of plotted finds (for example, bone, stone artefacts and so on) and lack of gaps in the stratigraphy shows that the occupation of PP5-6 after YTT was intense and continuous (Fig. 5 and Supplementary Videos 1–4). A 3D density analysis (see Methods) demonstrates that the amount of plotted finds remained continuous across the YTT event and increased in SADBS. The SADBS aggregate also contains a technological change, marked by the consistent use of heat-treated silcrete to produce stone tools and a commitment to advanced microlithic technology¹⁵. Not only did modern human populations on the south coast of South Africa survive the YTT event, but also the intensity of site use increased after this event, and a series of technological innovations appeared shortly after YTT and persisted for thousands of years¹⁵.

The southern African sub-region¹⁶ and south coast^{17,18} are candidate regions for housing the origin population of modern humans, and represent potential refuge zones during harsh climate cycles. It has previously been suggested that the Toba eruption caused a volcanic winter that devastated human populations^{5,19}, and a previously

The PP5-6 sediment stack profile

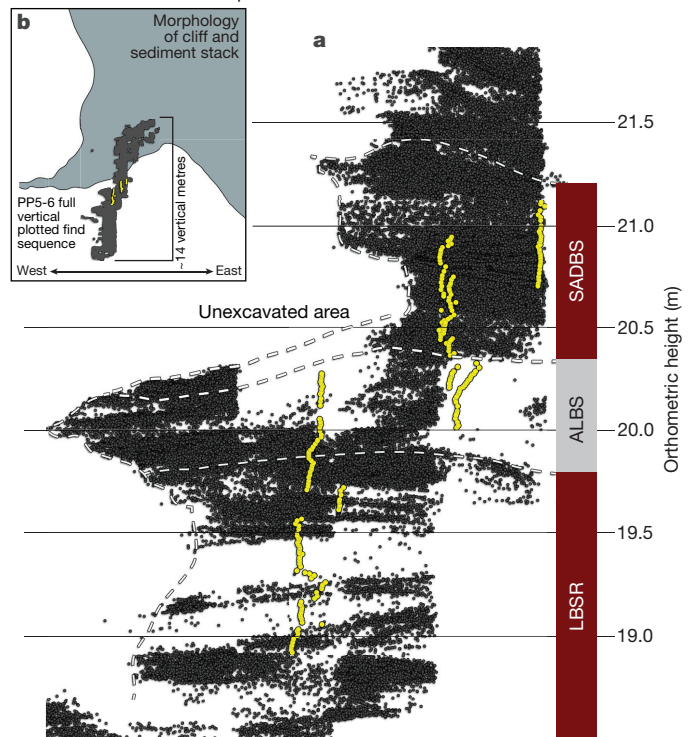


Figure 5 | The density of plotted finds across the upper LBSR, ALBS, and SADBS at PP5-6. This figure shows the distribution of plotted archaeological and faunal materials at PP5-6. Each dark grey point represents a single artefact, bone or shell. The yellow dots are sediment samples collected for analysis. The white dashes indicate the extent of the stratigraphic aggregate indicated as LBSR, ALBS or SADBS on the right. See Supplementary Videos 1–4 for 3D renderings of these data.

published climate model²⁰ has suggested that the YTT eruption had a decades-long negative effect on world terrestrial ecosystems. Our excavations at PP5-6 may, therefore, have intercepted a population in a coastal refuge during the hypothetical volcanic winter. Our results raise the question of whether the modern human population on the south coast of South Africa was the sole surviving population through a decade or more of volcanic winter, or whether populations elsewhere in Africa thrived through the YTT event and the ensuing MIS4 glacial phase. This question can now be tackled by applying the same suite of field and laboratory procedures^{8,21,22} used here to other archaeological sites in Africa.

Online Content Methods, along with any additional Extended Data display items and Source Data, are available in the online version of the paper; references unique to these sections appear only in the online paper.

Received 23 July 2015; accepted 2 February 2018.

Published online 12 March 2018.

- Williams, M. The ~73 ka Toba super-eruption and its impact: history of a debate. *Quat. Int.* **258**, 19–29 (2012).
- Storey, M., Roberts, R. G. & Saidin, M. Astronomically calibrated ⁴⁰Ar/³⁹Ar age for the Toba supereruption and global synchronization of late Quaternary records. *Proc. Natl Acad. Sci. USA* **109**, 18684–18688 (2012).
- Mark, D. F. *et al.* A high-precision ⁴⁰Ar/³⁹Ar age for the Young Toba Tuff and dating of ultra-distal tephra: forcing of Quaternary climate and implications for hominin occupation of India. *Quat. Geochronol.* **21**, 90–103 (2014).
- Marean, C. W. An evolutionary anthropological perspective on modern human origins. *Annu. Rev. Anthropol.* **44**, 533–556 (2015).
- Rampino, M. R. & Ambrose, S. H. in *Volcanic Hazards and Disasters in Human Antiquity* (eds McCoy, F. W. & Heiken, G.) 71–82 (Geological Society of America, 2000).
- Lane, C. S., Cullen, V. L., White, D., Bramham-Law, C. W. F. & Smith, V. C. Cryptotephra as a dating and correlation tool in archaeology. *J. Archaeol. Sci.* **42**, 42–50 (2014).
- Karkanis, P., Brown, K. S., Fisher, E. C., Jacobs, Z. & Marean, C. W. Interpreting human behavior from depositional rates and combustion features through the study of sedimentary microfossils at site Pinnacle Point 5-6, South Africa. *J. Hum. Evol.* **85**, 1–21 (2015).
- Oestmo, S. & Marean, C. W. in *Field Archaeology from Around the World* (eds Carver, M. B. *et al.*) 5955–5959 (Springer, 2015).
- Oestmo, S., Schoville, B. J., Wilkins, J. & Marean, C. W. A Middle Stone Age paleoscape near the Pinnacle Point caves, Vleesbaai, South Africa. *Quat. Int.* **350**, 147–168 (2014).
- Blockley, S. P. E. *et al.* A new and less destructive laboratory procedure for the physical separation of distal glass tephra shards from sediments. *Quat. Sci. Rev.* **24**, 1952–1960 (2005).
- Smith, V. C. *et al.* Geochemical fingerprinting of the widespread Toba tephra using biotite compositions. *Quat. Int.* **246**, 97–104 (2011).
- Lane, C. S., Chorn, B. T. & Johnson, T. C. Ash from the Toba supereruption in Lake Malawi shows no volcanic winter in East Africa at 75 ka. *Proc. Natl Acad. Sci. USA* **110**, 8025–8029 (2013).
- Westgate, J. A. *et al.* Tephrochronology of the Toba Tuffs: four primary glass populations define the 75-ka Youngest Toba Tuff, northern Sumatra, Indonesia. *J. Quat. Sci.* **28**, 772–776 (2013).
- Svensson, A. *et al.* Direct linking of Greenland and Antarctic ice cores at the Toba eruption (74 ka BP). *Clim. Past* **9**, 749–766 (2013).
- Brown, K. S. *et al.* An early and enduring advanced technology originating 71,000 years ago in South Africa. *Nature* **491**, 590–593 (2012).
- Henn, B. M. *et al.* Hunter-gatherer genomic diversity suggests a southern African origin for modern humans. *Proc. Natl Acad. Sci. USA* **108**, 5154–5162 (2011).
- Marean, C. W. Pinnacle Point Cave 13B (Western Cape Province, South Africa) in context: the Cape Floral kingdom, shellfish, and modern human origins. *J. Hum. Evol.* **59**, 425–443 (2010).
- Marean, C. W. *et al.* in *Fynbos: Ecology, Evolution, and Conservation of a Megadiverse Region* (eds Allsopp, N. *et al.*) 164–199 (Oxford Univ. Press, 2014).
- Ambrose, S. H. Late Pleistocene human population bottlenecks, volcanic winter, and differentiation of modern humans. *J. Hum. Evol.* **34**, 623–651 (1998).
- Robock, A. *et al.* Did the Toba volcanic eruption of ~74 ka B.P. produce widespread glaciation? *J. Geophys. Res. Atmos.* **114**, D10107 (2009).
- Fisher, E. C. *et al.* Technical considerations and methodology for creating high-resolution, color-corrected, and georectified photomosaics of stratigraphic sections at archaeological sites. *J. Archaeol. Sci.* **57**, 380–394 (2015).
- Bernatchez, J. A. & Marean, C. W. Total station archaeology and the use of digital photography. *SAA Archaeol. Rec.* **11**, 16–21 (2011).

Supplementary Information is available in the online version of the paper.

Acknowledgements This research was partially funded by the National Science Foundation (BCS-0524087 and BCS-1138073, C.W.M. and BCS-1460366, E.I.S. and C.W.M.), the Hyde Family Foundations (C.W.M.), the John Templeton Foundation (C.W.M.), the Institute of Human Origins at Arizona State University (C.W.M.), the Late Lessons from Early History program at ASU (C.W.M.), the ASU Strategic Initiative Fund, the Australian Research Council Discovery Project grant DP1092843 (Z.J.) and a Leverhulme Trust Early Career Fellowship (C.L.). S.O. thanks the American–Scandinavian Foundation and NORAM. A.C. was partially funded by an AAAS-Pacific Division, Alan E. Leviton Student Research Award and grants from the UNLV Department of Geoscience. We thank the MAPCRM staff for their assistance, T. Lachlan and Y. Jafari for help with OSL dating, the Dias Museum for field facilities and SAHRA and HWC for permits. The staff at the National Lacustrine Core Facility at the University of Minnesota (LacCore) provided a sample of Lake Malawi core for shard processing and analysis. M. Storey provided samples of YTT from Bukit Sapi, Malaysia. The opinions expressed in this publication are those of the author(s) and do not necessarily reflect the views of the funding agencies.

Author Contributions C.W.M. conceived and coordinated the study, and directed the fieldwork at PP5-6; S.O. and J.W. directed fieldwork at the Vleesbaai site; C.S.L. advised and assisted with cryptotephra methods and results; E.C.F. conducted the geographic information systems analysis, shard distribution analysis and co-directed the excavations; E.I.S., A.C., S.O., D.K. and J.W. collected samples for the cryptotephra study; E.I.S., R.J. and S.F. processed samples, identified sources and constructed the profile; J.A.H. conducted the Bayesian analysis of the geochemistry; M.R. analysed shards by electron probe microanalysis; N.C. helped to direct the excavations and collected many of the samples; J.A.H. provided the statistical model; P.K. studied the sedimentology and geology of the site and first discovered the shards; T.M. is an excavation permit co-holder and contributes to the palaeoenvironmental studies; and Z.J. conducted the OSL dating and Bayesian modelling of OSL ages. All authors contributed to the writing of the paper.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing interests. Readers are welcome to comment on the online version of the paper. Publisher's note: Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations. Correspondence and requests for materials should be addressed to E.I.S. (gene.smith@unlv.edu) for tephra study, and C.W.M. (curtis.marean@asu.edu) for archaeology.

Reviewer Information Nature thanks S. Blockley, R. Grun and the other anonymous reviewer(s) for their contribution to the peer review of this work.

METHODS

No statistical methods were used to predetermine sample size. The experiments were not randomized and investigators were not blinded to allocation during experiments and outcome assessment.

Cryptotephra extraction methods. We collected samples at PP5-6 and VBB by sampling along excavated sections every centimetre, and measured with total stations that provided millimetre accuracy. A total station is a modern surveying instrument that integrates an electronic theodolite with an electronic distance meter. A theodolite uses a movable telescope to measure angles in both the horizontal and vertical planes. For VBB samples, we combined every 5 cm of sample because of small sample size. We integrated location data into an ArcGIS 10.3 (<https://www.arcgis.com/>) geo-database with other data, including plotted finds, OSL samples, micromorphology samples and stratigraphic observations. The archaeological excavation methods used have previously been described^{8,21,22} as have the micromorphology methods⁷. In total, the samples and transects reported here from PP5-6 span the upper 1 m of LBSR, all of ALBS and the lower 60 cm of SADBS, where the shard count drops to zero. Natural variations in the slope and thickness of the sediments complicate linear 2D extrapolations of the sample locations within the sub-aggregates, so we calculated the relative height of each sample location above its nearest underlying sub-aggregate contact in each transect (Fig. 2); we also provide the raw shard counts per transect (Extended Data Figs 2, 3).

Because of the great travel distance, shards at PP5-6 and VBB are very rare and small. We label this type of tephra deposit, which is rarely found in complex archaeological sequences, extremely low abundance cryptotephra as a reflection of their low shard counts (<5 shards per gram) and small size (<50 µm). Extremely low abundance cryptotephra provide substantial processing and analytical challenges within cave and rock shelter sediments. Shards at VBB are fresh and provide excellent chemistry, but at PP5-6 they are altered and thus provide additional analytical challenges⁶. Separation of distal shards from sediment requires care and specialized techniques. In general, we followed a previously published method¹⁰ designed for processing lacustrine and organic sediments, but modified the procedure as described below to process extremely low abundance cryptotephra. We conducted preliminary processing at Oxford University under the guidance of C. Lane, and performed subsequent processing at the University of Nevada, Las Vegas.

Samples were weighed to 1 g and placed in 50-ml centrifuge tubes. A 10% HCl solution was added to dissolve carbonates. The samples were subsequently sieved at 80 and 20 µm. The 80–20-µm sieve portion was further processed using lithium metatungstate heavy liquid at densities of 1.95 g per cm³ and 2.55 g per cm³, to separate the glass shards and other low-density grains. Samples were centrifuged twice at each density for 15 min at 2,500 r.p.m. After cleaning, the separate (>1.95 but <2.55 g per cm³) that contained shards was mounted on a one-inch diameter epoxy round suitable for both major- and trace-element analyses. For VBB, only grains identified chemically as shards were counted (see 'Cryptotephra analytical methods'). For PP5-6, grains were first mounted on petrographic slides for identification and counting. This count is reported as shards per gram of dried sediment and plotted against vertical position in the sediment column to construct shard frequency diagrams. Petrographic slides (for PP5-6) and epoxy rounds (for VBB) were ground and polished for geochemical analysis.

Cryptotephra analytical methods. Extremely low abundance cryptotephra from PP5-6 occur as plate-like, bubble-wall and blocky glass shards (Extended Data Fig. 1). Shard shape was difficult to quantify because shards were identified after being mounted in epoxy. Using this type of mount, only 2D sections of the shards were available, which precluded detailed shape analysis. Most PP5-6 samples contained abundant biogenic silica that has shapes similar to shards, as previously noted²³. Biogenic silica was not present in the VBB samples but subangular-to-cusped quartz was common. For VBB, only grains verified by chemical analysis were positively identified as shards. For PP5-6, 49 tephra shards were counted that could be confidently distinguished from biogenic silica under the microscope. Most shards in PP5-6 sediments were either too small for analysis, were lost during sample polishing or showed signs of alteration. Some shards are 40 µm in size, and polished surfaces available for analysis were usually much smaller. Many grains originally thought to be shards were analysed as biogenic opal. Because of these difficulties, we were only able to obtain major-element analyses from 3 shards from PP5-6 in samples 48, 49 and 125, and trace elements from samples 48 and 49. From VBB, we obtained eight major-element and five trace-element analyses (Fig. 4, Extended Data Figs 5, 6 and Extended Data Tables 1, 2).

The <40-µm shards were analysed for major and minor elements using a JEOL JSX8900 SuperProbe with four wavelength dispersive spectrometers. Analytical conditions for electron microprobe analysis were 10 nA current at 15 kV accelerating voltage using a 10-µm spot size. Peak and background counting times were 30 and 10 s, respectively. Peak and background counting times for Na were 10 and 5 s. The alkali elements Na and K were counted on the first wavelength

dispersive spectrometer cycle to minimize potential element migration from beam damage. The rhyolite glass standard ATHO-G, which is part of the MPI-DING international standard set²⁴, was used as an internal standard for all analyses. Analytical error for most elements is less than ±0.2 weight per cent except SiO₂ (±0.83 weight per cent), Al₂O₃ (±0.25 weight per cent) and Na₂O (±0.57 weight per cent) (Supplementary Table 3).

Trace elements were obtained on a Thermo Scientific ICAP Q Quadrupole Inductively Coupled Plasma Mass Spectrometer (ICP-MS) at Michigan State University, in combination with a Photon Machines Analyte G2 193-nm excimer laser ablation system equipped with a 15 × 15-cm HelEx sample cell for solid sample microanalyses. Parameters used were 4.1 J per cm² laser fluence at 10 Hz (that is, 10 laser hits per second each delivering 4.1 J per cm²). The sample cell was fluxed in He, which was also used as the carrier gas to the ICP-MS injector. The ICP-MS was tuned using laser ablation of glass standard NIST 612 with the same fluence and frequency, performing a surface scan, aiming for highest intensities while keeping oxide production rate (ThO/Th) < 0.7% and double charged cations (¹³⁷Ba⁺⁺/¹³⁷Ba) < 3%. Laser pit size was adjusted based on sample size, but tuning and calibration were done at a 110-µm pit diameter on surface scans of NIST 612, USGS basalt glass standards and rock powder standards from the Geological Survey of Japan and the US Geological Survey (Supplementary Table 4). Backgrounds were assessed by collecting a gas blank after each standard and sample, and backgrounds were automatically subtracted from each analysis. Samples and standards were mounted on the same tray, enabling continuous analysis. Data for the Geological Survey of Japan standard JB1-a are provided in Supplementary Table 4.

For PP5-6 shard 49, we used the Cameca IMS 6f Secondary Ionization Mass Spectrometer (SIMS) at Arizona State University. Instrumental parameters for SIMS were primary beam: ¹⁶O⁻ at -12.5 kV, and current of 5 nA. Positive secondary ions were accelerated to 9,000 V into the mass spectrometer. Only secondary ions with 75 ± 20 eV excess kinetic energy were detected (conventional energy filtering). All ions were integrated for 5 s in each cycle of measurement, except for ³⁰Si, which was counted for 2 s in each cycle. Concentrations were determined by normalizing the secondary ion signal to that of ³⁰Si and referencing this ratio to the ion ratios previously determined on NIST 610 glass. The interference of BaO ions on Eu, NdO ions on Tb, NdO and SmO ions on Dy, SmO ions on Ho, NdO and SmO ions on Er, EuO ions on Tm, TbO ions on Lu, DyO ions on Hf and HoO ions on Ta were estimated using previously tabulated MO⁺/M⁺ values²⁵. Errors are two standard errors of the mean of 10 cycles of measurement. Errors on elements that required oxide corrections were estimated to be 20–30% larger than the error on the uncorrected ion ratio. LA-ICP-MS was used on shards with sufficient thickness to produce a significant analysis. For thin grains, SIMS was the preferred method of analysis. Supplementary Table 5 contains the errors for each element.

To monitor instrument accuracy and precision, the secondary standard rhyolite glass ATHO-G, part of the MPI-DING standard set²⁴, was analysed alongside glass shards on the electron microprobe. The average of these analyses and the per cent relative standard deviation (%RSD), which is used as an indication of instrument precision, are provided in Supplementary Table 3. The average of analyses for standard JB1-a used for LA-ICP-MS and the errors for each element for SIMS analysis can be found in Supplementary Tables 4, 5.

Cryptotephra source identification. Even moderate-sized eruptions can produce tephra that travels long distances²⁶, so we searched records of known volcanism beyond Africa to Antarctica, South America and Indonesia. We limited our search to Southern Hemisphere and equatorial-belt volcanoes because inter-hemispheric transport of volcanic ash is uncommon. To account for possible reworking of older cryptotephra into younger deposits, we widened our time window to include eruptions from over 700 ka to the present, including tephra found within the Siple A ice core from west Antarctica²⁷, the Rungwe volcanic province in Tanzania²⁸, the Turkana basin in Kenya and Ethiopia, distal ash from Turkana region eruptions in the Gulf of Aden^{29–31}, the database of obsidian from Kenya³², ash related to the Toba and Maninjau calderas in Indonesia^{33,34} and volcanoes in the southern Andes in Chile and Argentina^{35,36}. After this extensive search, we conclude that YTT is the source of the shards at both PP5-6 and VBB (see Supplementary Information for discussion).

Cryptotephra isochrons at PP5-6 and VBB. At PP5-6, the first appearance and highest concentration of shards occurs near the top of an aeolian sand unit (sub-aggregate Conrad Sand). This first peak is overlain by an approximately 10-cm-thick aeolian sand layer, which is free of shards, and then sealed by a series of successive intact hearth features also devoid of shards (in the lower Jocelyn sub-aggregate, Fig. 2). This confirms that wind-borne allochthonous material was able to enter the rock shelter at that time and that the first appearance of shards is a true isochron. The sand layer formed when the sea had regressed and the rock shelter opened onto a plain, providing a source for the wind-blown sand. A dune of similar age is found throughout Pinnacle Point and the shard input post-dates the beginning of this

dune activation. Such dunes move rapidly and can form thick sand layers within days. Above the hearth features, sub-aggregate Jocelyn shows a second increase in shards culminating within the overlying Erich sub-aggregate, both of which are located in sediments associated with trampled burnt remains⁷. These peaks and the overlying tail are interpreted to be reworkings of lateral exposures of the initial cryptotephra deposit by human trampling and continuous aeolian activity. We place the YTT isochron at the position of Conrad Sand (Fig. 2).

VBB has only a few shards in an aeolian deposit with extensive soil formation; placing an isochron there is challenging because the post-depositional migration of grains by mechanical and biological processes may mask the original position of the isochron. Bioturbation is especially common in sand dunes³⁷ and can cause grains to move down-section and even into units below the original depositional position. With the current results, we are not able to locate a more precise isochron in the VBB section. Although an isochron cannot be located, we do know that shards are found within and immediately above and below the SBCS unit dated to 74.7 ± 5.8 ka—a date that is nearly identical to the time of YTT (~ 74 ka)—and that the shards are an excellent match to the YTT chemistry.

Archaeological 3D density analysis methods. Stratigraphy at PP5-6 varies in both lateral and vertical morphology, which prevents simple 2D calculations of plotted-find point densities within these sediments. Therefore, we used ESRI ArcGIS 10.3 (<https://www.arcgis.com/>) to bin the plotted finds from the LBSR, ALBS and SADB (n = 117,975) into 125-cm³ cubes (5 cm × 5 cm × 5 cm) that were distributed evenly across the deposits in 3D space, using the South African National Grid (<http://www.ngi.gov.za/index.php/technical-information/geodesy-and-gps/datum-s-and-coordinate-systems>) (Lo. 23)) as input data. Counts of the points within each bin provided a standardized measure of 3D point density across the study deposits, at a resolution of 125 cm. Subsequent analyses were undertaken to assess the 3D density of points based on sub-aggregate and plotted find type (that is, stone artefact, bone, ochre and marine shell).

OSL methods. We have previously reported¹⁵ 25 ages for sediment samples collected from PP5-6. These ages have now been re-analysed, refined and increased in number in our current and ongoing study, providing dates with improved precision that supersede the previously reported ages¹⁵. No significant changes in ages were obtained (that is, all ages changed within their reported uncertainties); the major change was due to the updates associated with beta dose attenuation factors (using the latest published³⁸ values, which also affect the dose rate of the 'Nussli' standard used in our beta counting estimates, and a change in the error calculation on our beta dose rates). Full details of these changes and their implications have previously been published³⁹.

To date, 90 samples have been collected for OSL dating from the entire excavated profile of PP5-6 (Supplementary Table 6). Our overarching aim is to bracket each stratigraphic aggregate (for example, ALBS), and—where possible—to date many of the sub-aggregates (for example, Conrad Sand in ALBS), to provide high-resolution OSL dating to match the contextual control provided through total station plotting, stratigraphy, 3D GIS of plotted finds and lenses, and micro-morphology. This approach enables an assessment of temporal continuity (or lack thereof) between aggregates and estimate durations of specific lithic industries. Of particular use, the dense horizontal and vertical coverage of our ages can be input into a statistical age model for optimizing the resolution and precision of the chronology for the site. In this study, we use a Bayesian age model, using the OxCal platform (<https://c14.arch.ox.ac.uk/oxcal.html>), because this enables the relative stratigraphic information recorded at the site during excavation to be formally incorporated as a prior assumption in the model.

Experimental details are provided in a basic form: details of the chronology and associated methodological information have not yet been published, but the methods and handling of data are identical for every sample from PP5-6 and consistent with previously published methods^{40,41}.

OSL sample collection and preparation. We collected samples by driving a plastic tube (~ 1 cm in diameter, 20 cm long) into the cleaned section wall, or by using a cordless drill to loosen the samples (where necessary) and collecting the samples with a spoon at night using a red-filtered flashlight to ensure that sediments were collected from a single sub-aggregate. A sub-sample was also collected from the back of each hole for current soil moisture content measurements and for laboratory-based radioactivity measurements.

In the laboratory, the samples were opened under dim red illumination. Quartz grains of 180–212 μ m in diameter were selected for dating, after purifying the samples using HCl, H₂O₂ and solutions of sodium polytungstate at specific densities of 2.62 and 2.70 to separate the quartz grains from the feldspar and heavy minerals, respectively. The quartz grains were then etched in hydrofluoric acid to remove the external alpha-dosed layer (~ 10 μ m), and rinsed in HCl to dissolve any precipitated fluorides. The etched grains were sieved again to the lower sieve size (180 μ m).

OSL equivalent dose determination. Equivalent dose (D_e) values were estimated for individual sand-sized grains from all samples. The measurement conditions and analytical procedures most appropriate for individual grains from Pinnacle Point have previously been established⁴¹ and verified for the individual samples from PP5-6 through a range of dose recovery tests on representative samples from each of the stratigraphic aggregates. Measured dose:given dose ratios for these samples range between 0.97 ± 0.02 (sample 357332, aggregate OBS2) and 1.04 ± 0.04 (162516, RBSR), with overdispersion values of between 0% and $15 \pm 3\%$ and 1.01 ± 0.03 and overdispersion of $8 \pm 3\%$ for the sample from VBB (392138, SBCS). Overdispersion is the remaining amount of scatter after allowance is made for measurement uncertainties⁴². Because of this internal consistency, and because of the consistency between the VBB and PP5-6 samples and samples from other sites around Pinnacle Point, measurement and analytical procedures (including grain rejection, signal integration and preheat temperatures) were the same as previously used⁴¹. The equipment was also the same as that previously used^{15,41}. By using the same methods, equipment, procedures and operator in all of the single-grain OSL dating studies at Pinnacle Point, we can now place PP5-6 and VBB on the common timescale developed for Pinnacle Point, which enables improved comparisons of ages within and between sites.

Information about the number of grains measured and used, their overdispersion values calculated and the final $D_e \pm 1\sigma$ value for each sample is presented in Supplementary Table 6. The central age model⁴² was used to combine the individual single-grain D_e values of each sample to estimate the D_e for age calculation. The central age model assumes that the D_e values for all grains are centred on some average value of D (similar to the median) and the estimated standard error takes any overdispersion into account. Overdispersion values for these samples range between 0% (46800 and 46789 in LBSR and DBCS, respectively) and $35 \pm 4\%$ (418198 in ALBS) and, for samples from VBB, between $23 \pm 5\%$ and $30 \pm 6\%$ (Supplementary Table 6). Where overdispersion values are larger than expected, it is mostly due to one or two extreme outliers (see previously published plots¹⁵). All grains, regardless of whether they are extreme outliers or not, were included in the weighted mean value in case they form part of the natural variability of these samples. The weighted mean is insensitive to their inclusion.

OSL dose rate determination. The total dose rate consists of contributions from beta, gamma and cosmic radiation external to the grains, plus a small alpha dose rate due to the radioactive decay of U and Th inclusions inside sand-sized grains of quartz. To calculate the OSL ages, we have assumed that the measured radionuclide activities and dose rate have prevailed throughout the period of sample burial. An internal alpha dose rate of 0.036 ± 0.01 Gy ka⁻¹ has been assumed for all samples, based on measurements made on samples from the nearby site of Blombos Cave⁴³.

Bulk beta dose rates were measured directly by low-level beta counting of dried, homogenized and powdered sediment samples in the laboratory, using a GM-25-5 multi-counter system⁴⁴. The beta dose rate procedures used here have been shown to be both accurate and reproducible using a range of independent measurement techniques, with measurements performed in-house and by external laboratories⁴⁰. In this study, we have updated our previously published beta dose rate¹⁵, re-measured a selection of the samples, used the latest conversion factors and reassessed the errors associated with our beta dose rates, following the worked example³⁹ provided in Supplementary Information.

Gamma dose rates were measured at each sample location by *in situ* gamma spectrometry to take account of any spatial heterogeneity in the gamma radiation field within 30 cm of each OSL sample (as gamma rays can penetrate this distance through sediment and rock). Counts were collected for at least 60 min with a 1-inch Na(Tl) crystal. The detector was calibrated using the concrete blocks at Oxford⁴⁵ and the gamma dose rate was determined using the 'threshold' technique⁴⁶. This approach gives an estimate of the combined dose rate from gamma-ray emitters in the U and Th chains and from ⁴⁰K.

The cosmic-ray contribution was also taken into account, adjusting for site altitude, geomagnetic latitude, the density and thickness of rock and sediment overburden⁴⁷, and the $\cos^2\Phi$ zenith angle dependence of cosmic rays⁴⁸. We used the 3D GIS developed for PP5-6 as an aid to obtain accurate estimates of the various parameters that are input into the equation for estimating the cosmic-ray dose rate²¹ and considered the changes in dipline based on sedimentological observations to make informed decisions about the long-term patterns in rock overburden.

The beta, gamma and cosmic-ray dose rates were corrected for long-term water contents. The sandy layers and roofspall-rich layers are generally very dry, whereas those layers that contain substantial amounts of ash tend to be wetter. We assumed that this difference has always been the case. We used the current moisture contents as a guide and assigned a relative uncertainty of $\pm 25\%$ (at 1σ) to each estimate, unless it is a very low content, in which case an absolute error of $\pm 1\%$ was assigned (for example, $3 \pm 1\%$). At VBB, the current moisture content estimates

were generally very dry (<3%), so we have used a consistent value of $5 \pm 2\%$ for all samples. We think that this accommodates any likely variations over the average burial period of these samples. As a general rule of thumb for quartz grains, a 1% increase in water content will lead to an approximately 1% increase in age.

Statistical methods. *Cryptotephra*. Major element data from five volcanic fields (Maninjau, Andes Pat (Patagonia), Turkana, Antarctica and Rungwe) and six positively identified YTT distal samples (Malaysia, Lake Malawi, South China Sea, Bay of Bengal, Arabian Sea and India) were used as reference data for comparisons with unknown samples from PP5-6 and VBB. In this analysis, the six YTT distal datasets are combined into a single group within the major element analysis. Here we refer to the dataset of each volcanic field as a 'site'. We included 10 major elements from 791 samples from these sites (Supplementary Table 8a). Additionally, 25 trace elements from 333 samples from the following six sites were also included in the statistical analysis: Antarctica–Vostok, Kenya obsidian, Maninjau, Rungwe, Turkana and the YTT in Malaysia (Supplementary Table 8a). References for published data are included in the Supplementary Information. All data were log-transformed and standardized by subtracting their mean and dividing by the standard deviation. The element data were used to construct a Bayesian probability model, which assimilated the 10 variables associated with the major-element analysis and the 25-trace-element analysis (that is, two separate analyses were performed using the 10 major elements and the 25 trace elements). The full probability model describes the dependences among the data and parameters that produce posterior distributions: $P(\mu, \Sigma, p|C) \propto P(C|\mu, \Sigma)P(\mu)P(\Sigma)$. We assume the continuous trait data (major or trace element data), C , arise from a multivariate normal distribution. A mean vector μ and covariance matrix Σ describe the covariation among quantitative traits. The mean variable vector μ is assumed to vary by site (s) (for example, Antarctica, Rungwe). Thus, for the observation $i = 1, 2, \dots, n$, $C = 1, 2, \dots$, C quantitative traits and sites $s = 1, 2, \dots, 6$, $P(C|\mu, \Sigma)$ is given as

$$\begin{pmatrix} C_{1,i} \\ C_{2,i} \\ \vdots \\ C_{n,i} \end{pmatrix} \sim \text{Normal} \left(\begin{pmatrix} \mu_{1,s} \\ \mu_{2,s} \\ \vdots \\ \mu_{n,s} \end{pmatrix}, \Sigma \right)$$

We chose relatively non-informative conjugate priors for all quantities. An uninformative normal (Normal (0,10000) prior was specified for each component of μ and a relatively non-informative Wishart (R, k) prior⁴⁹ was chosen for the precision matrix ($\Omega = \Sigma^{-1}$), with $k = 10$ degrees of freedom and R as a 10×10 matrix.

The Bayesian model was implemented in OpenBUGS, a Bayesian simulation software⁵⁰. OpenBUGS uses Markov chain Monte Carlo (MCMC) routines to sample from the joint posterior distribution, enabling the generation of approximations of posterior distributions. Three parallel MCMC chains were run for 150,000 iterations and chains converged for all parameters and showed good mixing by 50,000 iterations. After thinning by 10 to reduce autocorrelation, a total of 30,000 iterations was used to calculate posterior statistics including the mean and 95% credible interval, defined by the 2.5th and 97.5th percentiles of each marginal posterior distribution.

We tested the predictive ability of our two models by randomly dividing our reference dataset of known volcanic eruptions into two subsets, 'training' data (90%) and 'validation' data (10%), to perform an out-of-sample cross validation. Each major- and trace-element dataset was subdivided into 'training' and 'validation' data subsets. For each dataset, 10% of the data was randomly assigned 'not assigned' (NA) for the variable 'site' (s). For every 'NA' sample, the model assigns the sample as belonging to each site with a given probability. A prior for the variable s is assigned as a categorical distribution with equal probability ($P_i = 0.1667$) for each of the six possible sites, using the equation $s_i \sim \text{categorical}(P_i)$. Drawing from the training data, the model thus assigns each 'unknown' sample a value for s . On completion of the out-of-sample cross validation, training and data subsets were recombined and used to analyse the archaeological samples from PP5-6 and VBB.

After removing the burn-in period ('burn-in' refers to the initial number of MCMC iterations before chain convergence), our model predicted the site of each 'unknown sample'. Accurate assignment of site was determined according to the highest probability of assignment. Values represent the proportion of MCMC iterations that assigned each sample to each respective site. For example, observation 18 (Supplementary Table 9, first entry) derives from the site Andes Pat. Drawing from distributions of major element data, the Bayesian model accurately assigned the sample with 100% probability. By contrast, the model incorrectly assigned observation 21 (Supplementary Table 9, second entry) from Andes Pat as Toba with 65% probability.

The results of the out-of-sample cross validation for the major element model suggest the model predicts at 92.4% accuracy (Supplementary Table 9); that is, the model accurately predicted the site for 'unknown' (samples known to the analyst but blind to the model) approximately 91.3% of the time. Six misidentified samples

were incorrectly assigned as Maninjau. Similarly, the model accurately differentiated sites using only trace element data with 79.4% accuracy (Supplementary Table 10). In contrast to the model using major element data, misidentified samples were likely to be assigned across multiple sites with equal probabilities rather than mis-assigned as a single site with high probability.

On assessing model performance, training and validation subsets were recombined and used to predict archaeological samples. Within the major element model, three samples derive from PP5-6 and seven samples derive from VBB. The major element model predicted both PP5-6 and VBB samples to belong to the Toba group with high probability (Extended Data Table 3). Similarly, the trace element model included two samples from PP5-6 and six samples from VBB, each predicted to belong to Toba with high probability (Extended Data Table 4).

OSL. Our single-grain OSL dates were put into a Bayesian statistical model on the OxCal platform (OxCal 4.2)^{51,52}. The dating estimates (Supplementary Table 6) were input as ' C_{cal} ' dates in calendar years before present, with an associated 1σ error. The OSL ages reported include both the random and systematic uncertainties. When ages with common systematic errors and random errors are combined, only the latter should be included. Therefore, we excluded all systematic errors⁵³ before constructing the Bayesian model; 'random-only' errors are shown in parentheses in the 'age' column of Supplementary Table 6.

To maximize the information that we obtained from the OSL dates for PP5-6, we used as prior information 10 of the 12 major stratigraphic aggregates. We did not include YBS because we only had one age for this unit; we included this age in YBSR. We also omitted the DBCS, which is a complex sedimentary unit that consists of sediment derived through debris flow from more than one stratigraphic aggregate (BBCSR and OBS2)⁷. We also used our knowledge regarding the relative stratigraphic position of each of the sub-aggregates that together make up the stratigraphic aggregates. We were therefore able to place our ages into sequences using the fine-scale stratigraphic information within each of the stratigraphic aggregates: we know, for example, that 'Conrad Cobble and Sand' in ALBS predates Conrad Sand and must therefore be older. Where we had more than one age estimate from a single sub-aggregate, we placed them into phases within the sequence, so that there were no assumptions about the relative order of the samples within the sub-aggregate.

We also used the outlier detection analysis⁵⁴ to assess the likelihood of each result being consistent within the constraints of the modelled sequence; this enables identification of possible erroneous ages. During the modelling process, the posterior outlier probability is calculated and the date down-weighted accordingly. For example, if the posterior probability is found to be 5%, the date is included in 95% of the model iterations, but if it is found to be 50% it is included in only 50% of model iterations. The outlier analysis detected seven significant outliers: specimen number 162660 (Aaron Sand and Roofspall), 418195 (Lwando Sand and Roofspall) and 110634 (Ludumo Sand and Roofspall) in LBSR, 418199 (Conrad Cobble and Sand) in ALBS, 357343 (Chris/Orfer) in OBS1, 357338 (Zuri Upper) in SGS and 162515 (Denise) in RBSR. This amounts to around 8%, which is slightly higher than statistical expectations (1 in 20 at 2σ).

The Bayesian age model for the samples from PP5-6 is shown in Fig. 2. The measured OSL ages are shown as black filled circles with their 1σ (68.2% probability) error bars, assuming a Gaussian distribution. The modelled ages are shown in orange as a continuous age range. These modelled age ranges are at 2σ (95.4% probability). The modelled probability distribution functions for the stratigraphic aggregate boundaries are also shown. Modelled age ranges are provided in Supplementary Table 6 and the stratigraphic aggregate boundaries are in Supplementary Table 7.

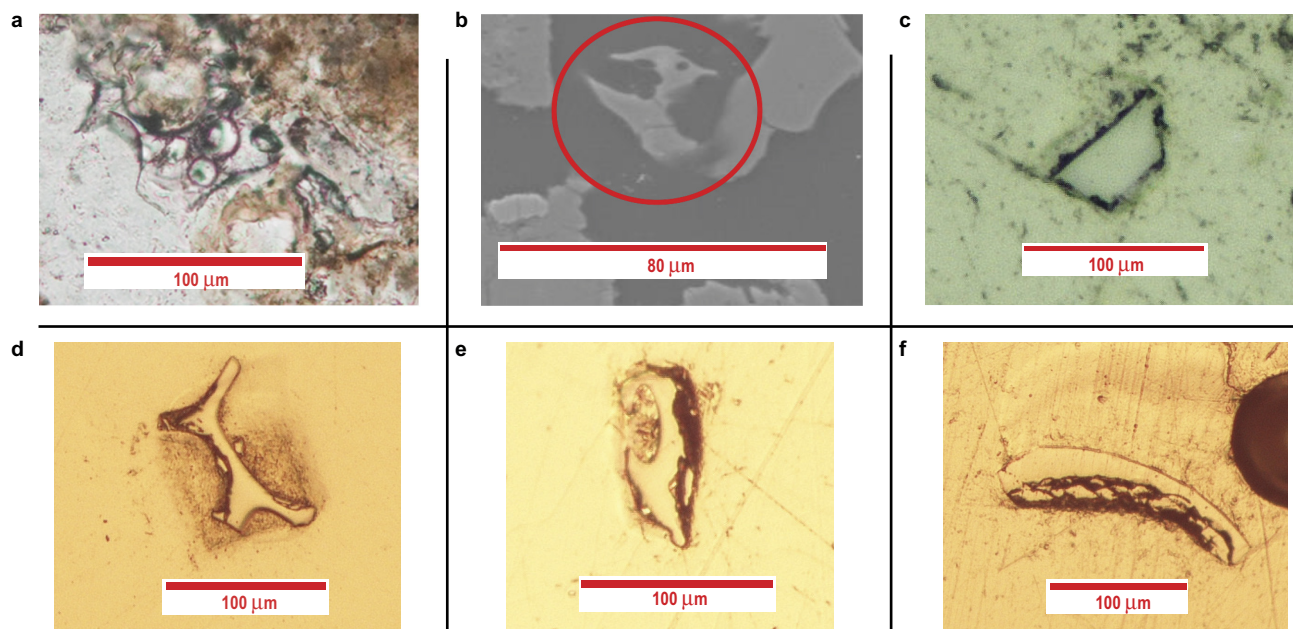
Sensitivity tests of the model have been conducted. The model was run four times using at least 50 million iterations and we confirmed that the modelled age estimates and posterior distribution functions for the stratigraphic boundaries were reproducible and that convergence values were high.

Code availability. The Bayesian age model coding is provided in the Supplementary Information.

Data availability. Most data generated or analysed during this study are included in the article and its Supplementary Information. All other data are available from the corresponding authors on reasonable request.

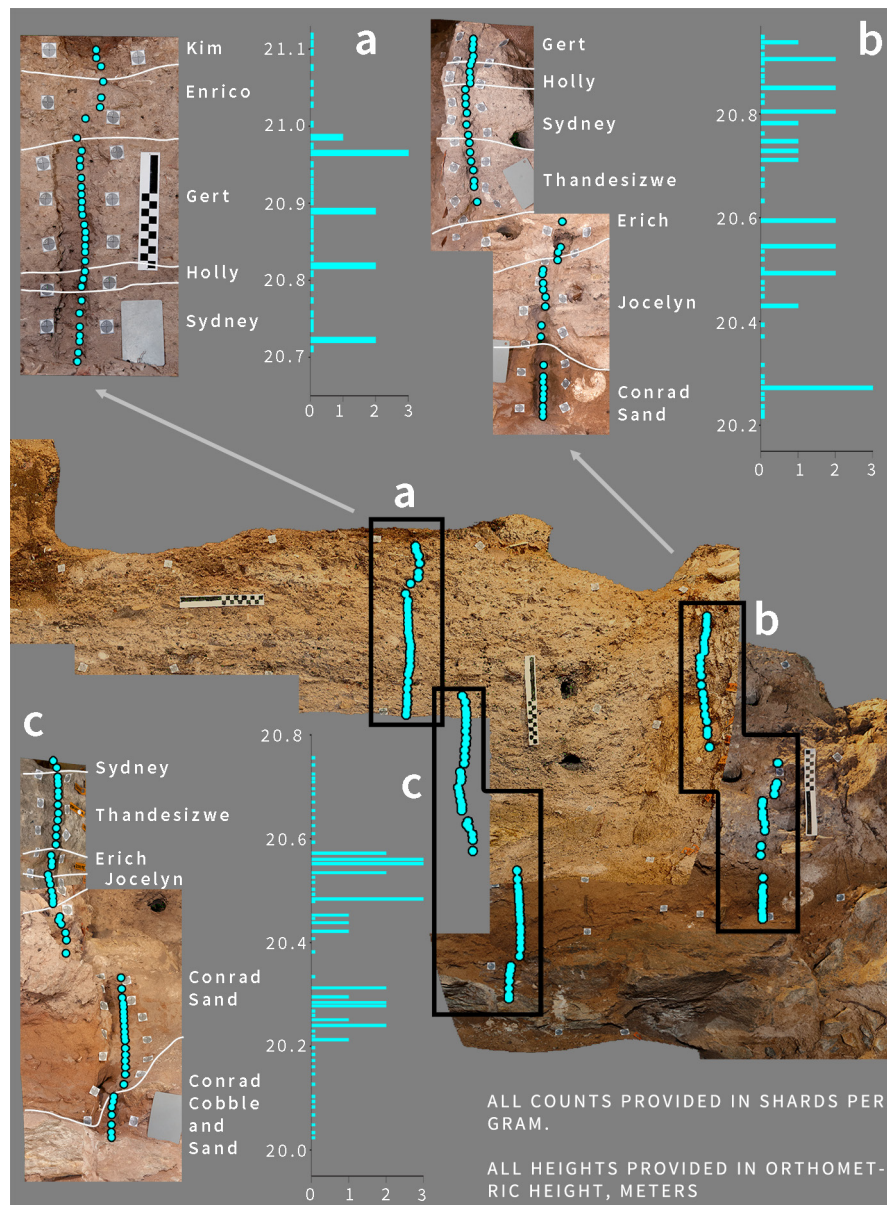
- Visser, M. P. E. *Detection of Middle to Late Holocene Icelandic Cryptotephra in the Netherlands: Tephra versus Biogenic Silica*. MSc thesis, Univ. Utrecht (2012).
- Jochum, K. P. *et al.* MPI-DING reference glasses for *in situ* microanalysis: new reference values for element concentrations and isotope ratios. *Geochim. Geophys. Res.* **7**, Q02008 (2006).
- Zinner, E. & Crozaz, G. A method for the quantitative measurement of rare earth elements in the ion microprobe. *Int. J. Mass Spectrom.* **69**, 17–38 (1986).
- Jensen, B. J. L. *et al.* Transatlantic distribution of the Alaskan White River Ash. *Geology* **42**, 875–878 (2014).

27. Dunbar, N. W. & Kurbatov, A. V. Tephrochronology of the Siple Dome ice core, West Antarctica: correlations and sources. *Quat. Sci. Rev.* **30**, 1602–1614 (2011).
28. Fontijn, K. *et al.* Holocene explosive eruptions in the Rungwe Volcanic Province, Tanzania. *J. Volcanol. Geotherm. Res.* **196**, 91–110 (2010).
29. Feakins, S. J., Brown, F. H. & deMenocal, P. B. Plio-Pleistocene microtephra in DSDP site 231, Gulf of Aden. *J. Afr. Earth Sci.* **48**, 341–352 (2007).
30. Brown, F. H., Haileab, B. & McDougall, I. Sequence of tuffs between the KBS Tuff and the Chari Tuff in the Turkana Basin, Kenya and Ethiopia. *J. Geol. Soc. London* **163**, 185–204 (2006).
31. Haileab, B. *Geochemistry, Geochronology and Tephrostratigraphy of Tephra from the Turkana Basin, Southern Ethiopia and Northern Kenya*. Ph.D. thesis, Univ. Utah (1995).
32. Brown, F. H., Nash, B. P., Fernandez, D. P., Merrick, H. V. & Thomas, R. J. Geochemical composition of source obsidians from Kenya. *J. Archaeol. Sci.* **40**, 3233–3251 (2013).
33. Chesner, C. A. & Luhr, J. F. A melt inclusion study of the Toba Tuffs, Sumatra, Indonesia. *J. Volcanol. Geotherm. Res.* **197**, 259–278 (2010).
34. Hashim, N. B. *Time Marker for the Late Pleistocene in Peninsular Malaysia: Study of the Volcanic Ash Deposits*. MSc thesis, Univ. Malaysia (2014).
35. Weller, D. J., Miranda, C. G., Moreno, P. I., Villa-Martínez, R. & Stern, C. R. Tephrochronology of the southernmost Andean Southern Volcanic Zone, Chile. *Bull. Volcanol.* **77**, 107 (2015).
36. Hildreth, W., Fierstein, J., Godoy, E., Drake, R. & Singer, B. The Puelche Volcanic Field: extensive Pleistocene rhyolite lava flows in the Andes of central Chile. *Rev. Geol. Chile* **26**, <http://dx.doi.org/10.4067/S0716-02081999000200008> (1999).
37. Ahlbrandt, T. S., Andrews, S. & Gwynne, D. T. Bioturbation in eolian deposits. *J. Sediment. Res.* **48**, 839–848 (1978).
38. Guérin, G., Mercier, N., Nathan, R., Adamiec, G. & Lefrais, Y. On the use of the infinite matrix assumption and associated concepts: a critical review. *Radiat. Meas.* **47**, 778–785 (2012).
39. Jacobs, Z. & Roberts, R. G. An improved single grain OSL chronology for the sedimentary deposits from Diepkloof Rockshelter, Western Cape, South Africa. *J. Archaeol. Sci.* **63**, 175–192 (2015).
40. Jacobs, Z., Roberts, R. G., Nespoulet, R., El Hajraoui, M. A. & Debénath, A. Single-grain OSL chronologies for Middle Palaeolithic deposits at El Mnasra and El Harhoura 2, Morocco: implications for Late Pleistocene human–environment interactions along the Atlantic coast of northwest Africa. *J. Hum. Evol.* **62**, 377–394 (2012).
41. Jacobs, Z. An OSL chronology for the sedimentary deposits from Pinnacle Point Cave 13B—a punctuated presence. *J. Hum. Evol.* **59**, 289–305 (2010).
42. Roberts, R. G., Galbraith, R. F., Olley, J. M., Yoshida, H. & Laslett, G. M. Optical dating of single and multiple grains of quartz from Jinmium Rock Shelter, Northern Australia: part II, results and implications. *Archaeometry* **41**, 365–395 (1999).
43. Jacobs, Z., Wintle, A. G. & Duller, G. A. T. Optical dating of dune sand from Blombos Cave, South Africa: I—multiple grain data. *J. Hum. Evol.* **44**, 599–612 (2003).
44. Bøtter-Jensen, L. & Mejdahl, V. Assessment of beta dose-rate using a GM multiscaler system. *Radiat. Meas.* **14**, 187–191 (1988).
45. Rhodes, E. J. & Schwenninger, J.-L. Dose rates and radioisotope concentrations in the concrete calibration blocks at Oxford. *Anc. TL* **25**, 5–8 (2007).
46. Mercier, N. & Falgueres, C. Field gamma dose-rate measurement with a NaI (TI) detector: re-evaluation of the ‘threshold’ technique. *Anc. TL* **25**, 1–4 (2007).
47. Prescott, J. R. & Hutton, J. T. Cosmic ray and gamma ray dosimetry for TL and ESR. *Int. J. Rad. Appl. Instrum. D* **14**, 223–227 (1988).
48. Smith, M. A., Prescott, J. R. & Head, J. N. Comparison of ^{14}C and luminescence chronologies at Pūritjarra rock shelter, central Australia. *Quat. Sci. Rev.* **16**, 299–320 (1997).
49. Gelman, A. *et al.* *Bayesian Data Analysis* (CRC, 2013).
50. Lunn, D., Spiegelhalter, D., Thomas, A. & Best, N. The BUGS project: evolution, critique and future directions. *Stat. Med.* **28**, 3049–3067 (2009).
51. Ramsey, C. B. Development of the radiocarbon calibration program. *Radiocarbon* **43**, 355–363 (2001).
52. Ramsey, C. B. Bayesian analysis of radiocarbon dates. *Radiocarbon* **51**, 337–360 (2009).
53. Rhodes, E. J. *et al.* Bayesian methods applied to the interpretation of multiple OSL dates: high precision sediment ages from Old Scatness Broch excavations, Shetland Isles. *Quat. Sci. Rev.* **22**, 1231–1244 (2003).
54. Ramsey, C. B. Dealing with outliers and offsets in radiocarbon dating. *Radiocarbon* **51**, 1023–1045 (2009).



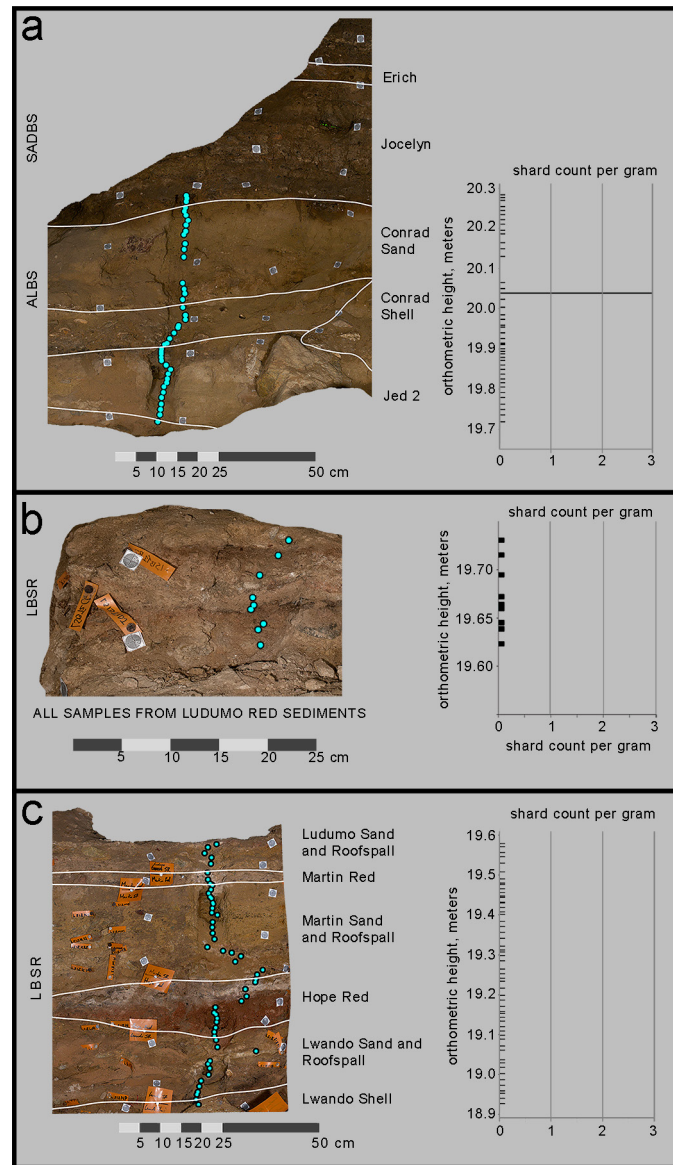
Extended Data Figure 1 | Examples of VBB and PP5-6 extremely low abundance cryptotephra. **a**, Two shard-like grains from PP5-6 in thin section (originally discovered by P.K.). **b**, Shard from PP5-6 sample 48

(scanning electron microscopy image). **c**, Shard from PP5-6 sample 125 (in thin section using plane-polarized light). **d–f**, Shards from VBB (from polished epoxy rounds using plane-polarized light).



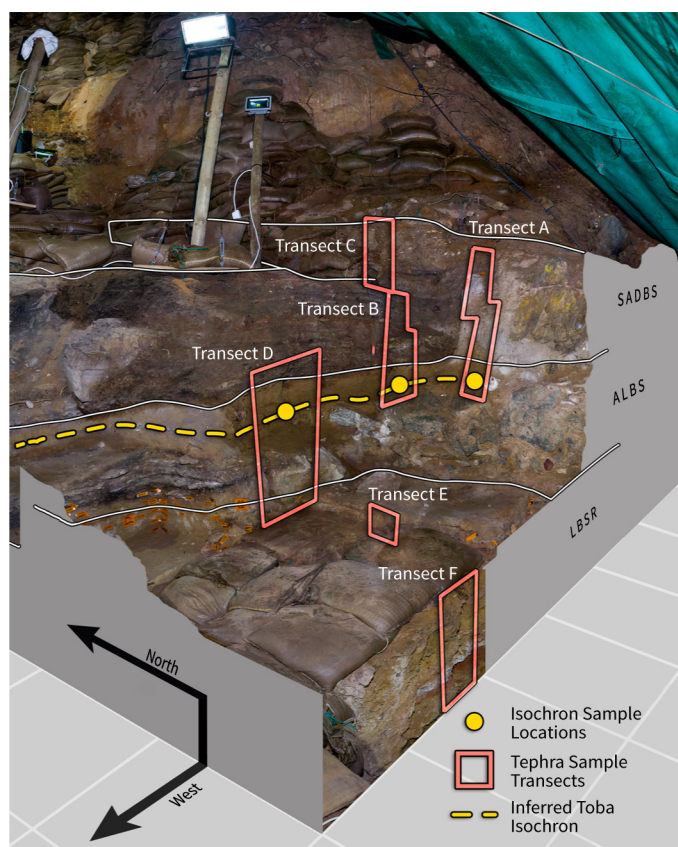
Extended Data Figure 2 | Individual sample transects on the sections and the shard counts from the transect. Sample transects are shown on the left and shard counts are shown on the right. **a**, Transect A. **b**, Transect B.

c, Transect C. The small bars showing shard counts of less than 1 indicate a sample with no shards. See Extended Data Fig. 4 for the overall location of the transects relative to one another.

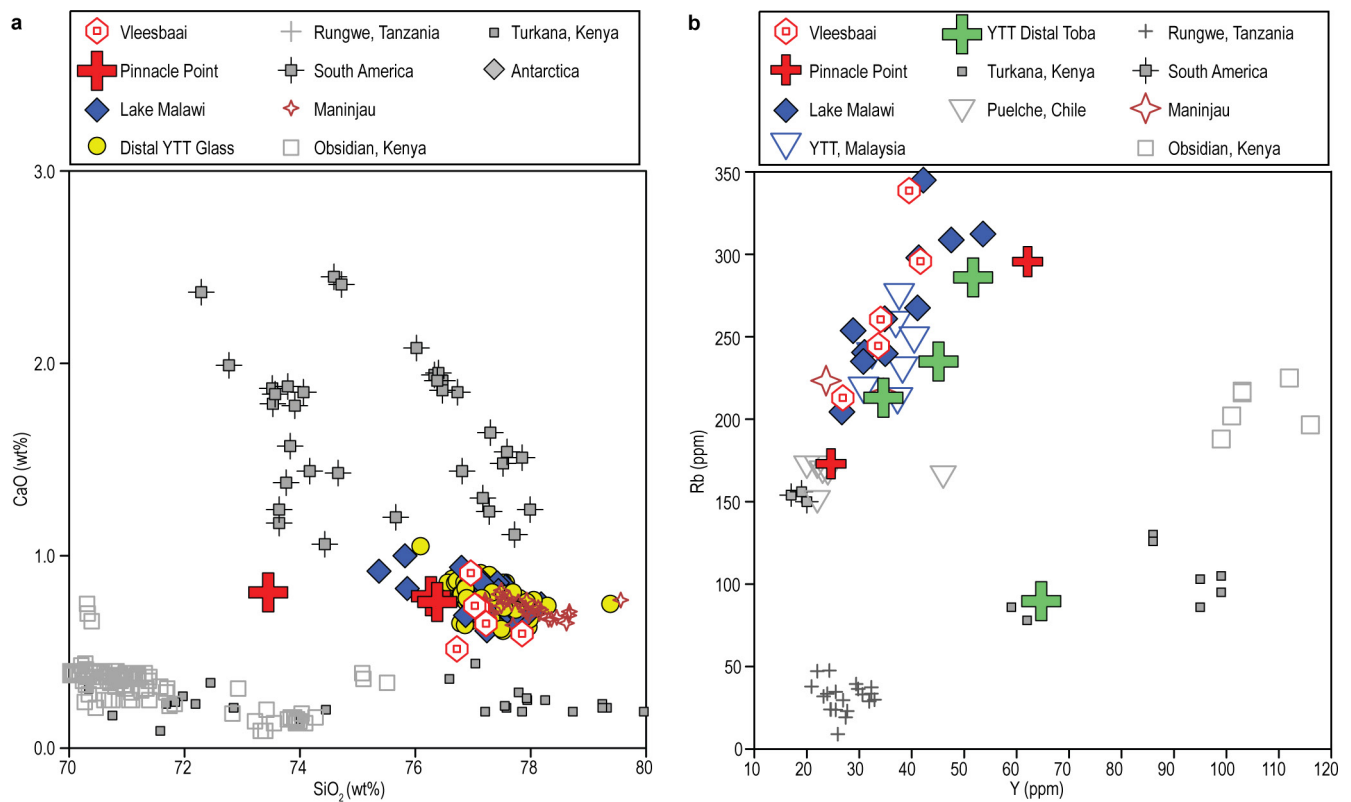


Extended Data Figure 3 | Individual transects on the sections and the shard counts per transect. a, Transect D. b, Transect E. c, Transect F. Transects are shown on the section (left), and shard counts per transect

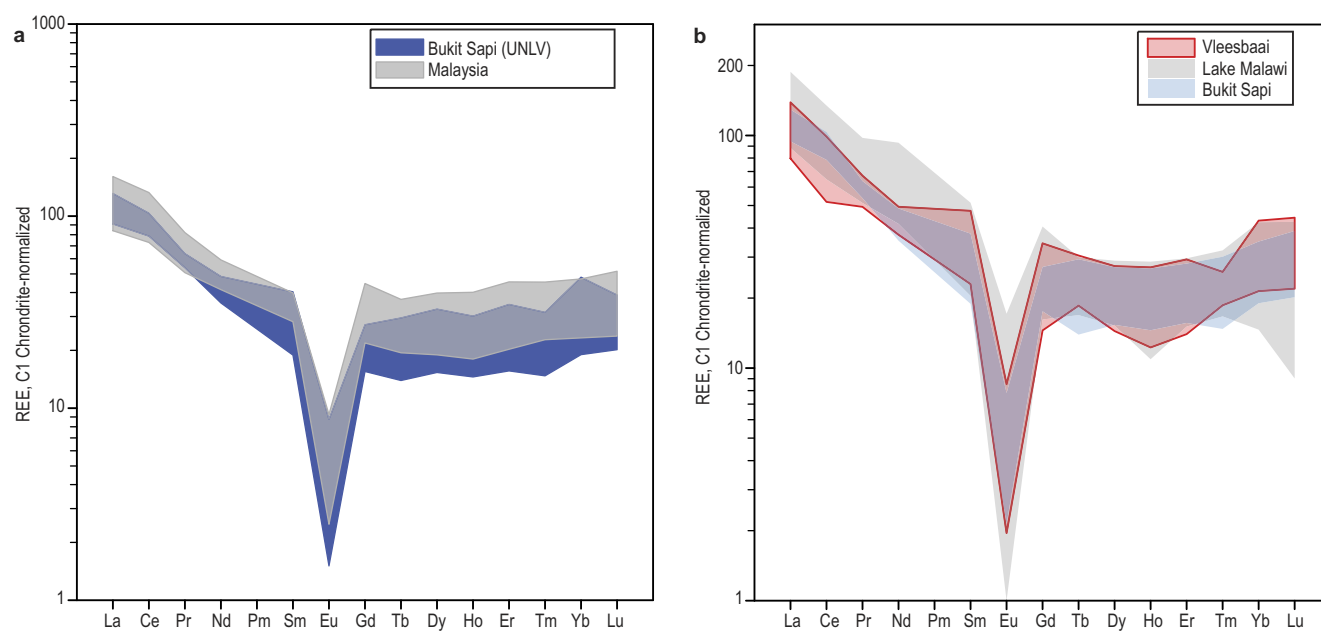
plotted (right). The small bars showing shard counts of less than 1 indicate a sample with no shards. See Extended Data Fig. 4 for the overall location of the transects relative to one another.



Extended Data Figure 4 | Panoramic photograph showing zones of contact between LBSR, ALBS and SADBS, and the location of shard sample transects. White lines indicate boundaries between stratigraphic aggregates and the yellow line indicates the YTT isochron.



Extended Data Figure 5 | Geochemical comparisons between the VBB and PP5-6 extremely low abundance cryptotephra and distal and proximal YTT. a, CaO versus SiO₂ (wt%). b, Rb versus Y (parts per million, p.p.m.). Note the change in symbols between a and b to separate YTT distal glass from Toba caldera and Malaysian samples.



Extended Data Figure 6 | Comparison of trace-element chemistry. **a**, Rare-earth element plot comparing new data for YTT from Bukit Sapi (Supplementary Table 1) to previously published data³⁴. **b**, Comparison of rare-earth element data for VBB, YTT from Lake Malawi and Bukit Sapi.

Extended Data Table 1 | Comparison of VBB, PP5-6, Lake Malawi, Bukit Sapi and Toba caldera shard analyses

| | VBB (8)* | STDEV** | PP5-6 (3) | STDEV | Lake Malawi (18) | STDEV | Bukit Sapi (31) | STDEV | Toba Caldera (118) | STDEV |
|--------------------------------|----------|---------|-----------|-------|------------------|-------|-----------------|-------|--------------------|-------|
| SiO ₂ | 77.20 | 0.34 | 75.37 | 1.66 | 77.24 | 1.79 | 77.46 | 1.16 | 77.22 | 0.70 |
| TiO ₂ | 0.05 | 0.08 | 0.21 | 0.10 | 0.05 | 0.04 | 0.05 | 0.05 | 0.06 | 0.06 |
| Al ₂ O ₃ | 12.41 | 0.23 | 13.57 | 1.88 | 12.41 | 0.31 | 12.33 | 0.29 | 12.54 | 0.39 |
| FeO | 0.81 | 0.07 | 1.29 | 0.39 | 0.84 | 0.14 | 0.90 | 0.34 | 0.85 | 0.24 |
| MnO | 0.07 | 0.03 | 0.09 | 0.03 | 0.07 | 0.08 | 0.06 | 0.08 | 0.07 | 0.09 |
| MgO | 0.06 | 0.03 | 0.08 | 0.07 | 0.05 | 0.06 | 0.06 | 0.06 | 0.05 | 0.04 |
| CaO | 0.70 | 0.12 | 0.79 | 0.03 | 0.77 | 0.19 | 0.78 | 0.09 | 0.78 | 0.21 |
| Na ₂ O | 3.23 | 0.29 | 2.90 | 0.54 | 2.95 | 0.37 | 3.05 | 0.19 | 3.10 | 0.34 |
| K ₂ O | 5.17 | 0.24 | 5.33 | 0.40 | 5.61 | 0.45 | 5.31 | 0.41 | 5.20 | 0.29 |
| P ₂ O ₅ | 0.01 | 0.01 | 0.05 | 0.04 | N.A. | N.A. | N.A. | N.A. | N.A. | N.A. |

| | MSU*** | MSU, ASU# | MSU (14) | MSU (12) | Toba Glass from Westgate et al. (13) | | | | | | | |
|-----------------------|--------|-----------|----------|----------|--------------------------------------|-------|-------|--------|-------------|-------------|-------------|--------------|
| Trace elements in ppm | | | | | | | | | | | | |
| Ga | 11.60 | 4.22 | N.A. | N.A. | 16.10 | 5.87 | 14.57 | 2.57 | N.A. | N.A. | N.A. | N.A. |
| Rb | 254 | 59.84 | 296 | 0.00 | 264 | 34.34 | 279 | 35.57 | 286 ± 34 | 253 ± 26 | 235 ± 31 | 213 ± 28 |
| Sr | 35.56 | 16.88 | 37.95 | 7.00 | 48.36 | 20.25 | 36.26 | 12.55 | 28.5 ± 4.9 | 56.2 ± 8.0 | 75.8 ± 13.5 | 108.0 ± 14.9 |
| Y | 31.64 | 10.02 | 43.35 | 26.52 | 37.51 | 6.73 | 36.42 | 7.41 | 64.7 ± 10.7 | 51.7 ± 10.3 | 45.1 ± 10.0 | 34.6 ± 5.5 |
| Zr | 79.34 | 11.19 | 278 | 46.67 | 81.90 | 19.35 | 81.37 | 11.08 | 108 ± 16 | 112 ± 16 | 122 ± 21 | 128 ± 18 |
| Nb | 16.70 | 2.50 | 51.60 | 25.60 | 18.23 | 6.19 | 17.16 | 2.29 | 22.5 ± 2.5 | 18.9 ± 2.2 | 17.8 ± 2.1 | 16.4 ± 1.7 |
| Cs | 10.19 | 3.22 | N.A. | N.A. | 11.09 | 2.38 | 11.31 | 2.42 | N.A. | N.A. | N.A. | N.A. |
| Ba | 231 | 21.43 | 241 | 4.24 | 356 | 65.38 | 283 | 183.28 | 88 ± 23 | 382 ± 54 | 737 ± 145 | 1142 ± 104 |
| La | 24.35 | 4.93 | 76.60 | 10.75 | 29.00 | 6.08 | 25.18 | 3.08 | 32.3 ± 4.3 | 36.9 ± 5.4 | 43.9 ± 7.3 | 60.7 ± 11.2 |
| Ce | 45.95 | 10.23 | 147 | 26.87 | 59.09 | 13.29 | 54.83 | 4.36 | 62 ± 6.5 | 67.4 ± 7.6 | 74.3 ± 8.6 | 99.2 ± 13.2 |
| Pr | 5.23 | 0.86 | 16.45 | 3.32 | 6.32 | 0.92 | 5.53 | 0.32 | 7.42 ± 1.05 | 7.70 ± 1.21 | 8.55 ± 1.56 | 10.12 ± 1.45 |
| Nd | 18.97 | 2.78 | 50.05 | 3.04 | 22.45 | 3.39 | 19.00 | 1.73 | 28.1 ± 4.5 | 28.4 ± 6.2 | 29.7 ± 6.1 | 34.4 ± 5.7 |
| Sm | 4.37 | 1.85 | 7.40 | 1.56 | 4.77 | 1.45 | 4.23 | 1.02 | 7.43 ± 1.63 | 6.66 ± 1.52 | 6.52 ± 1.91 | 6.36 ± 2.92 |
| Eu | 0.26 | 0.13 | 0.69 | 0.13 | 0.43 | 0.19 | 0.29 | 0.12 | 0.32 ± 0.27 | 0.44 ± 0.26 | 0.59 ± 0.37 | 0.74 ± 0.32 |
| Gd | 4.15 | 1.62 | 4.51 | 0.00 | 4.85 | 0.87 | 4.18 | 0.78 | 7.95 ± 2.25 | 6.91 ± 2.43 | 6.61 ± 2.40 | 5.59 ± 1.98 |
| Tb | 0.73 | 0.24 | 0.82 | 0.00 | 0.76 | 0.15 | 0.77 | 0.20 | 1.42 ± 0.32 | 1.21 ± 0.35 | 1.07 ± 0.34 | 0.88 ± 0.27 |
| Dy | 4.64 | 1.60 | 3.94 | 0.37 | 5.33 | 1.12 | 5.35 | 1.32 | 9.64 ± 2.06 | 7.91 ± 1.90 | 7.37 ± 2.20 | 5.53 ± 1.49 |
| Ho | 0.93 | 0.38 | 0.83 | 0.04 | 1.11 | 0.24 | 1.15 | 0.26 | 2.17 ± 0.46 | 1.82 ± 0.48 | 1.62 ± 0.47 | 1.26 ± 0.37 |
| Er | 3.09 | 1.25 | 2.67 | 0.38 | 3.34 | 0.85 | 3.70 | 0.90 | 6.99 ± 1.40 | 5.78 ± 1.67 | 5.15 ± 1.66 | 3.93 ± 1.00 |
| Tm | 0.48 | 0.20 | 0.47 | 0.05 | 0.62 | 0.19 | 0.57 | 0.14 | 1.18 ± 0.28 | 0.97 ± 0.27 | 0.84 ± 0.29 | 0.65 ± 0.21 |
| Yb | 4.34 | 1.88 | 2.71 | 0.01 | 4.13 | 1.28 | 4.58 | 1.33 | 7.96 ± 1.35 | 6.30 ± 1.13 | 5.74 ± 1.44 | 4.26 ± 0.93 |
| Lu | 0.71 | 0.26 | 0.40 | 0.00 | 0.63 | 0.20 | 0.65 | 0.15 | 1.28 ± 0.27 | 1.01 ± 0.28 | 0.92 ± 0.28 | 0.69 ± 0.18 |
| Hf | 3.09 | 0.78 | 5.82 | 2.72 | 2.76 | 0.83 | 3.15 | 0.39 | 5.47 ± 1.22 | 5.02 ± 1.06 | 5.16 ± 1.19 | 5.08 ± 1.17 |
| Ta | 1.96 | 0.53 | 2.13 | 0.25 | 1.77 | 0.48 | 2.08 | 0.35 | 3.48 ± 0.57 | 2.72 ± 0.43 | 2.45 ± 0.47 | 1.95 ± 0.34 |
| Pb | 34.20 | 3.86 | 27.30 | 3.11 | 32.50 | 4.42 | 38.35 | 2.90 | 59.9 ± 18.6 | 60.7 ± 12.5 | 58.3 ± 16.5 | 59.8 ± 12.8 |
| Th | 30.86 | 9.21 | 14.10 | 1.84 | 29.76 | 7.35 | 32.58 | 4.71 | 58 ± 8.9 | 48.8 ± 7.9 | 45.5 ± 8.3 | 40.5 ± 5.9 |
| U | 5.72 | 1.27 | 3.80 | 0.42 | 4.56 | 1.32 | 5.35 | 0.99 | 9.97 ± 1.44 | 7.55 ± 0.99 | 6.34 ± 1.09 | 5.30 ± 0.66 |

Numbers in parentheses are sample sizes. N.A., not applicable.

*UNLV EPMA, analysis by electron probe microanalysis at the University of Nevada Las Vegas.

**STDEV, two standard deviations.

***MSU, analysis by LA-ICP-MS at Michigan State University.

#ASU, analysis by SIMS at Arizona State University.

Extended Data Table 2 | Major and minor trace element chemistry for VBB and PP5-6

Major Elements in wt% by EPMA

| | VBB VBB6 | VBB VBB6-2 | VBB VBB16-1 | VBB VBB16-2 | VBB VBB9 | VBB VBB8 | VBB VBB-7-2 | VBB VBB12-2 | PP5-6 PP125 | PP5-6 PP49 | PP5-6 PP48 |
|--------------------------------|-------------|---------------|----------------|----------------|-------------|-------------|----------------|----------------|----------------|---------------|---------------|
| SiO ₂ | 76.96 | 77.41 | 77.09 | 77.03 | 77.23 | 77.85 | 76.72 | 77.32 | 76.27 | 76.38 | 73.45 |
| TiO ₂ | 0.02 | 0.07 | 0 | 0.21 | 0.01 | 0 | 0.11 | 0 | 0.12 | 0.18 | 0.32 |
| Al ₂ O ₃ | 12.4 | 12.35 | 12.49 | 12.82 | 12 | 12.27 | 12.46 | 12.49 | 12.68 | 12.3 | 15.73 |
| FeO | 0.85 | 0.85 | 0.88 | 0.81 | 0.79 | 0.66 | 0.77 | 0.89 | 0.85 | 1.61 | 1.41 |
| MnO | 0.08 | 0.06 | 0.07 | 0.01 | 0.05 | 0.1 | 0.06 | 0.11 | 0.11 | 0.05 | 0.11 |
| MgO | 0.05 | 0.06 | 0.04 | 0.08 | 0.11 | 0.04 | 0.03 | 0.05 | 0.04 | 0.04 | 0.16 |
| CaO | 0.91 | 0.82 | 0.73 | 0.74 | 0.65 | 0.6 | 0.52 | 0.66 | 0.79 | 0.76 | 0.81 |
| Na ₂ O | 3.18 | 3.03 | 3.54 | 2.91 | 3.19 | 3.5 | 3.62 | 2.89 | 3.43 | 2.92 | 2.36 |
| K ₂ O | 4.98 | 5.07 | 4.92 | 5.21 | 5.3 | 4.99 | 5.66 | 5.19 | 5.51 | 5.6 | 4.87 |
| P ₂ O ₅ | 0.04 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0.07 | 0 | 0.07 |
| F | 0.2 | 0.13 | 0.03 | 0 | 0 | 0 | 0 | 0.26 | 0 | 0.01 | 0.61 |
| Cl | 0.43 | 0.2 | 0.18 | 0.19 | 0.18 | 0.08 | 0.01 | 0.15 | 0.14 | 0.15 | 0.17 |
| Original Total | 96.43 | 97.33 | 94.84 | 95.41 | 95.14 | 94.93 | 100.12 | 96.68 | 94.04 | 96.11 | 92.95 |

Trace Elements in ppm by LA-ICP-MS, PP49 by SIMS

| | | | | | | | | | | | |
|----|-------|-------|-------|-------|-------|-------|------|------|------|-------|-------|
| Ga | 13.22 | 17.87 | 9.44 | 12.88 | 10.9 | 5.26 | N.A. | N.A. | N.A. | N.A. | N.A. |
| Rb | 260.5 | 169.3 | 295.8 | 212.9 | 338.6 | 244.5 | N.A. | N.A. | N.A. | 296.0 | 296.0 |
| Sr | 52.94 | 51.54 | 15.64 | 47.59 | 23.52 | 22.13 | N.A. | N.A. | N.A. | 33 | 42.9 |
| Y | 34.07 | 14.12 | 41.66 | 26.86 | 39.5 | 33.62 | N.A. | N.A. | N.A. | 62.1 | 24.6 |
| Zr | 95.31 | 74.72 | 77.89 | 82.77 | 83.8 | 61.55 | N.A. | N.A. | N.A. | 245 | 311 |
| Nb | 16.79 | 16.8 | 19.55 | 14.16 | 19.32 | 13.59 | N.A. | N.A. | N.A. | 69.7 | 33.5 |
| Cs | 9.9 | 5.65 | 13.78 | 8.1 | 13.91 | 9.78 | N.A. | N.A. | N.A. | N.A. | N.A. |
| Ba | 522.4 | 213.3 | 51.7 | 468.5 | 60.4 | 67.2 | N.A. | N.A. | N.A. | 238.0 | 244.0 |
| La | 32.89 | 22.37 | 23.05 | 27 | 21.87 | 18.89 | N.A. | N.A. | N.A. | 84.2 | 69 |
| Ce | 60.6 | 37.65 | 47.84 | 51.91 | 45.94 | 31.77 | N.A. | N.A. | N.A. | 166 | 128 |
| Pr | 6.23 | 3.93 | 5.14 | 5.72 | 5.8 | 4.58 | N.A. | N.A. | N.A. | 18.8 | 14.1 |
| Nd | 22.56 | 15.18 | 21.69 | 18.98 | 18.31 | 17.11 | N.A. | N.A. | N.A. | 52.2 | 47.9 |
| Sm | 4.8 | 1.75 | 7.02 | 3.66 | 5.6 | 3.4 | N.A. | N.A. | N.A. | 8.5 | 6.3 |
| Eu | 0.48 | 0.29 | 0.13 | 0.3 | 0.25 | 0.11 | N.A. | N.A. | N.A. | 0.6 | 0.78 |
| Gd | 4.43 | 2.25 | 4.76 | 3.75 | 6.84 | 2.89 | N.A. | N.A. | N.A. | N.A. | 4.51 |
| Tb | 0.67 | 0.35 | 0.73 | 0.7 | 0.82 | 1.1 | N.A. | N.A. | N.A. | N.A. | 0.82 |
| Dy | 5.11 | 2.66 | 6.09 | 3.68 | 6.75 | 3.55 | N.A. | N.A. | N.A. | 4.2 | 3.67 |
| Ho | 0.98 | 0.38 | 1.1 | 0.97 | 1.48 | 0.67 | N.A. | N.A. | N.A. | 0.8 | 0.85 |
| Er | 2.81 | 1.36 | 4.69 | 2.24 | 4.29 | 3.12 | N.A. | N.A. | N.A. | 2.4 | 2.94 |
| Tm | 0.6 | 0.1 | 0.64 | 0.5 | 0.58 | 0.46 | N.A. | N.A. | N.A. | 0.5 | 0.43 |
| Yb | 4.62 | 1.43 | 5.57 | 3.45 | 6.93 | 4.01 | N.A. | N.A. | N.A. | 2.7 | 2.72 |
| Lu | 0.54 | 0.34 | 0.77 | 0.67 | 0.84 | 1.09 | N.A. | N.A. | N.A. | 0.4 | 0.4 |
| Hf | 3.32 | 3.76 | 3.37 | 3.42 | 3.14 | 1.55 | N.A. | N.A. | N.A. | 3.9 | 7.74 |
| Ta | 1.66 | 2.11 | 2.59 | 1.42 | 2.54 | 1.42 | N.A. | N.A. | N.A. | 2.3 | 1.95 |
| Pb | 34.76 | 33.36 | 38.82 | 31.68 | 37.99 | 28.61 | N.A. | N.A. | N.A. | 29.5 | 25.1 |
| Th | 32.19 | 15.87 | 39.08 | 27.29 | 41.54 | 29.16 | N.A. | N.A. | N.A. | 12.8 | 15.4 |
| U | 4.81 | 5.74 | 6.84 | 4.47 | 7.63 | 4.82 | N.A. | N.A. | N.A. | 4.1 | 3.5 |

Trace elements in p.p.m. by LA-ICP-MS.

Extended Data Table 3 | The posterior estimates of site for each archaeological sample in the major elements model

| Site | Model Predicted Site | | | | | |
|-----------|----------------------|------------|----------|--------|------|---------|
| | Andes Pat | Antarctica | Maninjau | Rungwe | Toba | Turkana |
| PP5-6 | 0.00 | 0.00 | 0.00 | 0.00 | 1.00 | 0.00 |
| PP5-6 | 0.00 | 0.00 | 0.01 | 0.00 | 0.99 | 0.00 |
| PP5-6 | 0.00 | 0.00 | 0.00 | 0.00 | 0.97 | 0.03 |
| Vleesbaai | 0.00 | 0.00 | 0.00 | 0.00 | 0.98 | 0.02 |
| Vleesbaai | 0.00 | 0.00 | 0.02 | 0.00 | 0.98 | 0.00 |
| Vleesbaai | 0.00 | 0.00 | 0.00 | 0.00 | 0.98 | 0.02 |
| Vleesbaai | 0.00 | 0.00 | 0.00 | 0.01 | 0.99 | 0.00 |
| Vleesbaai | 0.00 | 0.00 | 0.00 | 0.00 | 0.99 | 0.01 |
| Vleesbaai | 0.00 | 0.00 | 0.00 | 0.00 | 0.99 | 0.01 |
| Vleesbaai | 0.00 | 0.00 | 0.00 | 0.00 | 0.99 | 0.01 |

Values represent the proportion of MCMC iterations that assigned each of the six samples as belonging to each of the six sites.

Extended Data Table 4 | The posterior estimates of site for each archaeological sample in the trace elements model

| Site | Model Predicted Site | | | | | |
|-----------|----------------------|-------|----------|--------|------|---------|
| | Antarctica | Kenya | Maninjau | Rungwe | Toba | Turkana |
| PP5-6 | 0.00 | 0.00 | 0.00 | 0.00 | 0.99 | 0.01 |
| PP5-6 | 0.00 | 0.01 | 0.00 | 0.00 | 0.99 | 0.00 |
| Vleesbaai | 0.00 | 0.00 | 0.00 | 0.00 | 1.00 | 0.00 |
| Vleesbaai | 0.00 | 0.00 | 0.00 | 0.00 | 1.00 | 0.00 |
| Vleesbaai | 0.00 | 0.00 | 0.00 | 0.00 | 1.00 | 0.00 |
| Vleesbaai | 0.00 | 0.01 | 0.00 | 0.01 | 0.97 | 0.01 |
| Vleesbaai | 0.00 | 0.00 | 0.00 | 0.00 | 1.00 | 0.00 |
| Vleesbaai | 0.00 | 0.00 | 0.00 | 0.00 | 1.00 | 0.00 |

Values represent the proportion of MCMC iterations that assigned each of the six samples as belonging to each of the six sites.

Carbon dioxide addition to coral reef waters suppresses net community calcification

Rebecca Albright^{1,2}, Yuichiro Takeshita^{1,3}, David A. Kowek¹, Aaron Ninokawa⁴, Kennedy Wolfe⁵, Tanya Rivlin^{6,7}, Yana Nebuchina¹, Jordan Young⁴ & Ken Caldeira¹

Coral reefs feed millions of people worldwide, provide coastal protection and generate billions of dollars annually in tourism revenue¹. The underlying architecture of a reef is a biogenic carbonate structure that accretes over many years of active biomineralization by calcifying organisms, including corals and algae². Ocean acidification poses a chronic threat to coral reefs by reducing the saturation state of the aragonite mineral of which coral skeletons are primarily composed, and lowering the concentration of carbonate ions required to maintain the carbonate reef. Reduced calcification, coupled with increased bioerosion and dissolution³, may drive reefs into a state of net loss this century⁴. Our ability to predict changes in ecosystem function and associated services ultimately hinges on our understanding of community- and ecosystem-scale responses. Past research has primarily focused on the responses of individual species rather than evaluating more complex, community-level responses. Here we use an *in situ* carbon dioxide enrichment experiment to quantify the net calcification response of a coral reef flat to acidification. We present an estimate of community-scale calcification sensitivity to ocean acidification that is, to our knowledge, the first to be based on a controlled experiment in the natural environment. This estimate provides evidence that near-future reductions in the aragonite saturation state will compromise the ecosystem function of coral reefs.

Oceanic uptake of atmospheric carbon dioxide (CO₂) is causing measurable declines in surface ocean pH, the concentration of carbonate ions and Ω_{arag} ^{5,6}. The aragonite saturation state (Ω_{arag}) of tropical-Pacific surface water is estimated to have decreased from values of about 4.5 in pre-industrial times^{7,8} to about 3.8 by 1995⁹, and is expected to continue declining to approximately 3.0 by the middle of the century and 2.3 by the end of the century⁹. To test whether ocean acidification is already impairing the net community calcification (NCC) (calcification minus dissolution) of coral reefs, an alkalization experiment was previously conducted, in which the Ω_{arag} value of seawater flowing over a reef flat was restored to near pre-industrial levels¹⁰. Using a novel dual tracer regression method, the authors estimated changes in NCC (that is, alkalinity uptake¹¹) in response to increased Ω_{arag} and showed that present-day calcification rates are depressed compared to the values expected for pre-industrial conditions. In other words, ocean acidification may already be impairing coral reef growth. Here we return to the site of the original alkalization study and apply the dual tracer regression method to quantify the calcification response to CO₂ enrichment. In doing so, we test the hypothesis that near-future reductions in Ω_{arag} will significantly impair NCC at the community scale.

In contrast to laboratory settings, CO₂ enrichment in the natural environment enables the investigation of the calcification response of a coral reef to high CO₂ under natural abiotic conditions (such as temperature, light, flow and nutrients) and accounting for species

interactions. We used CO₂ gas to increase the CO₂ content of seawater flowing over a patch of coral reef to achieve Ω_{arag} levels that are projected to occur this century. A non-reactive dye tracer (Rhodamine WT, hereafter referred to as the dye) was added to the seawater along with CO₂, and a dual tracer regression technique¹⁰ was used to estimate changes in NCC in response to CO₂ enrichment. This technique uses the change in ratios between an active tracer (alkalinity) and a passive tracer (dye) to assess the change in alkalinity uptake (NCC) owing to experimental conditions. Changes in the active tracer (alkalinity) result from mixing, dilution and biological activity (that is, NCC), whereas changes in the passive tracer (dye) are due solely to mixing and dilution. By comparing the alkalinity–dye ratios before (upstream of the study site) and after (downstream) the water mass interacts with the reef, we were able to quantify the change in NCC resulting from CO₂ enrichment (see Methods and Extended Data Fig. 1).

Our study was conducted on a 400-m² patch of unconfined reef flat at One Tree Island in the southern Great Barrier Reef (Fig. 1). This section of the reef flat is a well-developed, mixed-reef community characterized by approximately 15% live coral and 26% crustose coralline algae (see Methods). The experiment was conducted once per day, over 30 days from September to October 2016. Prior to low tide each day, a 15-m³ tank was deployed adjacent to the study site and filled with ambient seawater. On 20 of the days, CO₂ was bubbled into the tank to lower the tank pH from ambient levels (approximately 8.1) to approximately 6.1 (Extended Data Fig. 2). The pH content of the header tank was intentionally low to account for dilution upon release to the reef flat. The relevant pH is the pH as the water enters our experimental zone. In the study site, the seawater pH was lowered by just 0.143 ± 0.007 units (mean \pm s.e.m.), from a background value of 8.126 ± 0.013 to a treatment level of 7.983 ± 0.015 . About 4 g Rhodamine WT was also added to the tank and the solution was mixed thoroughly. On 10 of the days, observations were made when dye but no CO₂ was added (hereafter referred to as control days), to test whether the dye addition had unexpected effects, and to characterize background variability in the study area. The resulting solution was pumped onto the reef flat at a constant rate of approximately 31 s^{-1} for 60 min starting at the predicted time of low tide; the resulting plume was allowed to flow over the reef flat. After the 60-min pumping period, discrete water samples were taken at pre-defined sampling locations along the length of two parallel transects that defined the borders of the study area (along the upstream and downstream edges of the reef flat; Fig. 1). Samples were analysed for total alkalinity, Rhodamine WT and pH, as described in the Methods (Supplementary Table 1).

Paired alkalinity–dye measurements from all days were analysed using a multivariate regression approach to calculate alkalinity–dye ratios (slopes) and mean background alkalinities (y intercepts) of the upstream and downstream transects, while simultaneously accounting for natural spatial and temporal variability (Methods, Extended Data

¹Department of Global Ecology, Carnegie Institution for Science, Stanford, California 94305, USA. ²California Academy of Sciences, San Francisco, California 94118, USA. ³Monterey Bay Aquarium Research Institute, Moss Landing, California 95039, USA. ⁴UC Davis Bodega Marine Lab, Bodega Bay, California 94923, USA. ⁵School of Medical Sciences, The University of Sydney, Sydney, New South Wales 2006, Australia. ⁶The Interuniversity Institute for Marine Sciences, The H. Steinitz Marine Biology Laboratory, The Hebrew University of Jerusalem, Coral Beach, Eilat, 8810300, Israel. ⁷The Fredy and Nadine Herrman Institute of Earth Sciences, The Hebrew University of Jerusalem, Edmond J. Safra Campus, Jerusalem, 9190401, Israel.

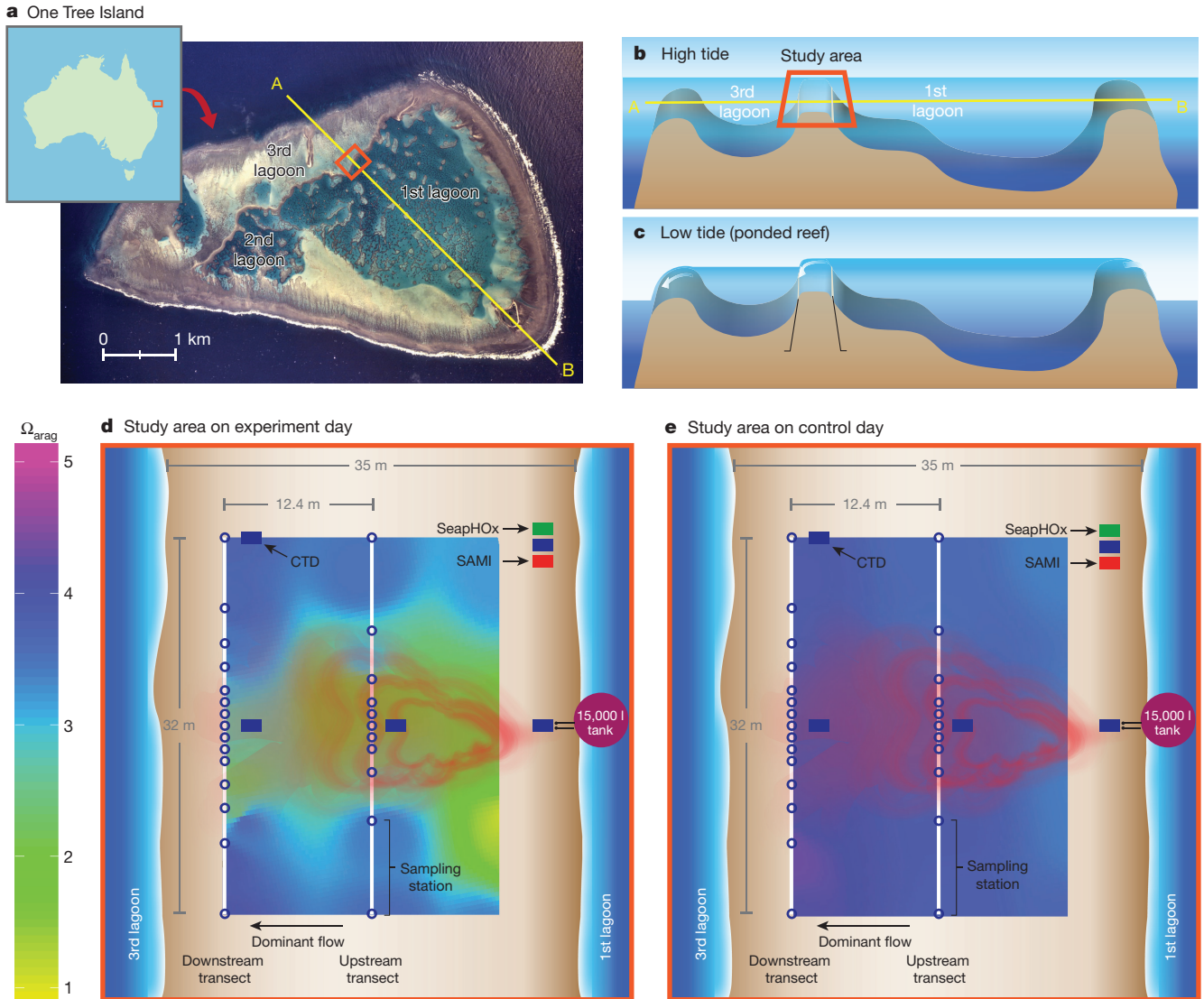


Figure 1 | Study site. **a**, Map of Australia and an aerial photograph of One Tree Island with the study area denoted in orange. Use of the photograph was permitted under an educational license from the University of Sydney. **b**, **c**, Cross sections along the yellow line are shown for high (**b**) and low (**c**) tides. The reef wall containing the study site separates the first lagoon and the third lagoon. The first lagoon sits approximately 30 cm higher than the third lagoon, and at low tide this results in gravity-driven, unidirectional flow from the first lagoon, over the reef flat study area, and into the third lagoon. This unidirectional flow across the study site persists for approximately 60 min after peak low tide, catering to the experimental

design depicted in **d** and **e**. **d**, **e**, Schematic of the study area (to scale) indicating positioning of the transects and the sampling locations (blue circles). Heat maps show Ω_{arag} within the study area for a representative experiment day ($n = 20$ independent experiments, 29 September 2016 shown here) (**d**) and control day ($n = 10$ independent experiments, 30 September 2016 shown here) (**e**). CTD, conductivity, temperature and depth sensing system; SAMI, submersible autonomous moored instrument monitoring p_{CO_2} ; SeapHOx, conductivity, temperature, depth, pH, oxygen sensor.

design depicted in **d** and **e**. **d**, **e**, Schematic of the study area (to scale) indicating positioning of the transects and the sampling locations (blue circles). Heat maps show Ω_{arag} within the study area for a representative experiment day ($n = 20$ independent experiments, 29 September 2016 shown here) (**d**) and control day ($n = 10$ independent experiments, 30 September 2016 shown here) (**e**). CTD, conductivity, temperature and depth sensing system; SAMI, submersible autonomous moored instrument monitoring p_{CO_2} ; SeapHOx, conductivity, temperature, depth, pH, oxygen sensor.

Fig. 3). The percentage change in net calcification, ΔG , resulting from CO_2 enrichment can be calculated as:

$$\Delta G = \frac{G_{\text{decrease}}}{G_{\text{background}}} \quad (1)$$

in which G_{decrease} is the change in NCC, inside the plume, resulting from CO_2 enrichment in mmol s^{-1} , and $G_{\text{background}}$ is the background calcification rate in mmol s^{-1} (that is, the calcification rate without added CO_2). G_{decrease} and $G_{\text{background}}$ can be calculated as:

$$G_{\text{decrease}} = P_{\text{dye}}(r_{\text{up}} - r_{\text{down}}) \quad (2)$$

$$G_{\text{background}} = F(a_{\text{up}} - a_{\text{down}}) \quad (3)$$

in which P_{dye} is the pumping rate of the dye in g s^{-1} , r_{up} and r_{down} are the alkalinity-dye ratios (slopes) for the upstream and down-

stream transects, respectively, in $\mu\text{mol kg}^{-1} \text{ppb}^{-1}$ or mmol g^{-1} , F is the volumetric flow rate in $\text{m}^3 \text{s}^{-1}$, and a_{up} and a_{down} are the mean background alkalinities (that is, the y intercepts) of the upstream and downstream transects, respectively, in mmol m^{-3} (see Methods).

CO_2 enrichment lowered the Ω_{arag} in the plume (defined as the area containing 95% of the dye) by 0.79 ± 0.03 units (mean \pm s.e.m.), from a background value of 3.70 ± 0.08 to 2.91 ± 0.08 (Fig. 2, Extended Data Fig. 4). The 0.79 unit reduction in Ω_{arag} inside the plume decreased NCC by 34%, from a background value of $3.80 \pm 0.24 \mu\text{mol m}^{-2} \text{s}^{-1}$ to $2.52 \pm 0.26 \mu\text{mol m}^{-2} \text{s}^{-1}$ (Fig. 2). On control days, Ω_{arag} averaged 3.76 ± 0.11 , and there was no significant difference in Ω_{arag} inside the plume compared to background conditions (Fig. 2, Extended Data Fig. 4). On these days, there was no significant difference between NCC inside the plume ($3.84 \pm 0.41 \mu\text{mol m}^{-2} \text{s}^{-1}$) compared to background values ($3.91 \pm 0.44 \mu\text{mol m}^{-2} \text{s}^{-1}$, Fig. 2, Extended Data Fig. 4). Paired

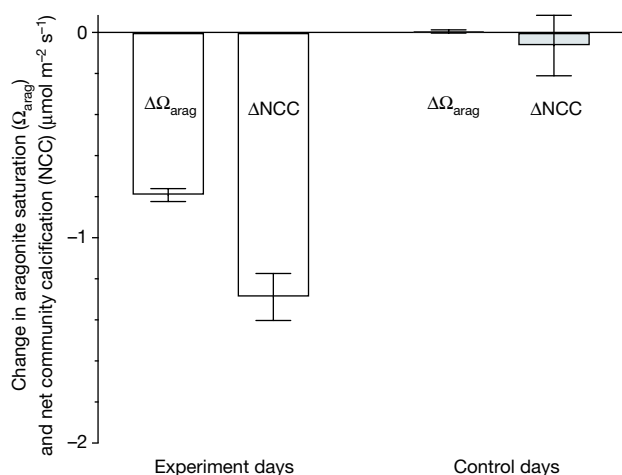


Figure 2 | Mean change in Ω_{arag} and NCC. Change (mean \pm s.e.m.) in Ω_{arag} and NCC inside the plume compared to background conditions for experiment days ($n = 20$ independent experiments) and control days ($n = 10$ independent experiments). Error bars reflect underlying natural variability (that is, day-to-day and hour-to-hour), as Ω_{arag} and NCC varied based on time of day and light availability. Results by day are presented in Extended Data Fig. 6.

t -tests were used to compare NCC inside the plume with background values on experiment ($t_{19} = 11.26$, $P < 0.0001$) and control ($t_9 = 0.43$, $P > 0.05$) days, rejecting the null hypothesis that reductions in Ω_{arag} projected to occur this century do not impair NCC. Results by day are presented in Fig. 3.

Previous estimates of the calcification sensitivity of corals to Ω_{arag} on the basis of laboratory studies range from 15% per unit Ω (ref. 12) to 28% (ref. 13), although a uniform response is not observed for all species. Care should be taken, however, when comparing calcification relationships derived from coral studies to mixed-reef communities such as that of our study site. The high sensitivity of calcification to CO_2 enrichment observed in this study may, in part, be due to a high abundance of crustose coralline algae (26%), which secrete high-magnesium calcite, a form of calcium carbonate that is more soluble than aragonite. On comparison of these results with previously published work¹⁰, it appears that the sensitivity of NCC to Ω_{arag} may increase as Ω_{arag} decreases. Notably, in both our alkalization¹⁰ and acidification studies we did not observe a relationship between background calcification rates and changes in NCC (that is, the change in NCC was not proportional to background NCC, Extended Data Fig. 5), which may suggest that seawater chemistry primarily influences dissolution as opposed to gross calcification³. Further studies investigating the relative effects of ocean acidification on gross calcification versus dissolution may help constrain the mechanism behind reductions in NCC.

CO_2 enrichment experiments without enclosure are common in terrestrial studies¹⁴ and have been used to characterize responses in seagrasses and marine plants¹⁵. While other *in situ* enrichment studies (for example, nutrients¹⁶) have been conducted in reef environments, to our knowledge, this study represents the first CO_2 enrichment experiment of an unconfined coral reef system, representing a key step towards scaling our understanding of the effects of ocean acidification from individual organisms to community responses¹⁷. Taken alongside the results of other study systems (for example, mesocosms¹⁸, flumes¹⁹, CO_2 seeps^{20–22} and areas with naturally high CO_2 variability²³), there is growing evidence that ocean acidification will cause profound, ecosystem-wide changes in coral reefs, including loss of three-dimensional framework and associated biodiversity and ecosystem services.

Given the recent acceleration in global reef decline from acute impacts such as mass bleaching^{24,25}, acidification-induced reductions in reef growth and population replenishment will severely compromise coral reef recovery and could threaten their survival²⁶. Reef declines

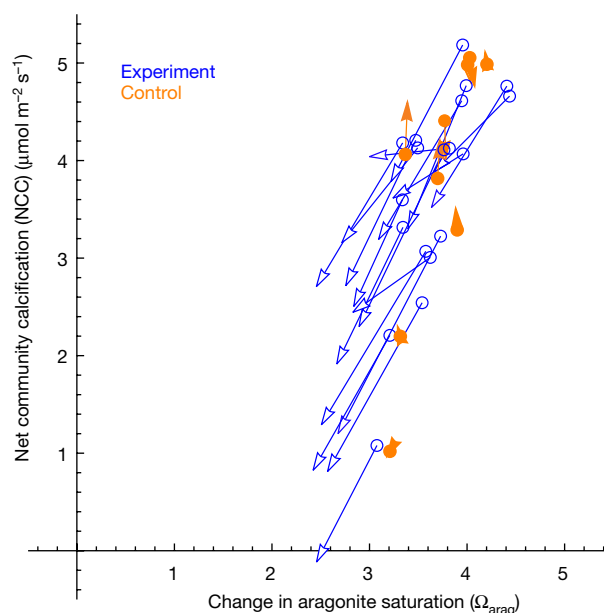


Figure 3 | Change in Ω_{arag} and NCC by day. Change in Ω_{arag} and the associated change in NCC, for experiment days ($n = 20$ independent experiments) and control days ($n = 10$ independent experiments). Circles represent starting (that is, background) conditions and arrows represent ending (that is, in-plume) conditions. The short arrows on control days indicate uncertainty due to unexplained variability.

may have a wide range of possible effects on coastal human communities including reduced food, income and well-being, as well as longer-term effects such as increased vulnerability as coral reefs become less able to protect coastal areas from storms and waves^{27–29}. Our challenge is to understand how ecosystems are likely to respond to future changes and how changes in natural resources are likely to affect human livelihoods, and then to act on this information in a way that maximizes desired outcomes²⁶. Ultimately, only the reduction of atmospheric CO_2 levels will address the challenges of ocean acidification. By examining the effects of excess CO_2 on an unconfined coral reef community, our results add to the mounting body of evidence that without deep and rapid reductions in emissions, the persistence of coral reefs is at risk^{24,30}.

Online Content Methods, along with any additional Extended Data display items and Source Data, are available in the online version of the paper; references unique to these sections appear only in the online paper.

Received 1 August 2017; accepted 2 February 2018.

Published online 14 March 2018.

- Spalding, M. *et al.* Mapping the global value and distribution of coral reef tourism. *Mar. Policy* **82**, 104–113 (2017).
- Andersson, A. J. & Gledhill, D. Ocean acidification and coral reefs: effects on breakdown, dissolution, and net ecosystem calcification. *Ann. Rev. Mar. Sci.* **5**, 321–348 (2013).
- Eyre, B. D., Andersson, A. J. & Cyronak, T. Benthic coral reef calcium carbonate dissolution in an acidifying ocean. *Nat. Clim. Chang.* **4**, 969–976 (2014).
- Silverman, J., Lazar, B., Cao, L., Caldeira, K. & Erez, J. Coral reefs may start dissolving when atmospheric CO_2 doubles. *Geophys. Res. Lett.* **36**, (2009).
- Bates, N. *et al.* A time-series view of changing ocean chemistry due to ocean uptake of anthropogenic CO_2 and ocean acidification. *Oceanography (Wash. D.C.)* **27**, 126–141 (2014).
- Le Quéré, C. *et al.* Global carbon budget 2014. *Earth Syst. Sci. Data* **7**, 47–85 (2015).
- Cao, L. & Caldeira, K. Atmospheric CO_2 stabilization and ocean acidification. *Geophys. Res. Lett.* **35**, (2008).
- Ricke, K. L., Orr, J. C., Schneider, K. & Caldeira, K. Risks to coral reefs from ocean carbonate chemistry changes in recent earth system model projections. *Environ. Res. Lett.* **8**, 034003 (2013).
- Feely, R. A., Doney, S. C. & Cooley, S. R. Ocean acidification: present conditions and future changes in a high- CO_2 world. *Oceanography (Wash. D.C.)* **22**, 36–47 (2009).

10. Albright, R. *et al.* Reversal of ocean acidification enhances net coral reef calcification. *Nature* **531**, 362–365 (2016).
11. Smith, S. V. & Key, G. S. Carbon dioxide and metabolism in marine environments. *Limnol. Oceanogr.* **20**, 493–495 (1975).
12. Chan, N. C. & Connolly, S. R. Sensitivity of coral calcification to ocean acidification: a meta-analysis. *Glob. Chang. Biol.* **19**, 282–290 (2013).
13. Langdon, C. & Atkinson, M. J. Effect of elevated pCO₂ on photosynthesis and calcification of corals and interactions with seasonal change in temperature/irradiance and nutrient enrichment. *J. Geophys. Res.* **110**, (2005).
14. Norby, R. J. & Zak, D. R. Ecological lessons from free-air CO₂ enrichment (FACE) experiments. *Annu. Rev. Ecol. Syst.* **42**, 181–203 (2011).
15. Arnold, T. *et al.* Ocean acidification and the loss of phenolic substances in marine plants. *PLoS ONE* **7**, e35107 (2012).
16. Koop, K. *et al.* ENCORE: the effect of nutrient enrichment on coral reefs. Synthesis of results and conclusions. *Mar. Pollut. Bull.* **42**, 91–120 (2001).
17. Andersson, A. *et al.* Understanding ocean acidification impacts on organismal to ecological scales. *Oceanography (Wash. D.C.)* **28**, 16–27 (2015).
18. Dove, S. G. *et al.* Future reef decalcification under a business-as-usual CO₂ emission scenario. *Proc. Natl Acad. Sci. USA* **110**, 15342–15347 (2013).
19. Kline, D. I. *et al.* A short-term *in situ* CO₂ enrichment experiment on Heron Island (GBR). *Sci. Rep.* **2**, 413 (2012).
20. Fabricius, K. E. *et al.* Losers and winners in coral reefs acclimatized to elevated carbon dioxide concentrations. *Nat. Clim. Chang.* **1**, 165–169 (2011).
21. Enochs, I. C. *et al.* Shift from coral to macroalgae dominance on a volcanically acidified reef. *Nat. Clim. Chang.* **5**, 1083–1088 (2015).
22. Inoue, S., Kayanne, H., Yamamoto, S. & Kurihara, H. Spatial community shift from hard to soft corals in acidified water. *Nat. Clim. Chang.* **3**, 683–687 (2013).
23. Price, N. N., Martz, T. R., Brainard, R. E. & Smith, J. E. Diel variability in seawater pH relates to calcification and benthic community structure on coral reefs. *PLoS ONE* **7**, e43843 (2012).
24. Hughes, T. P. *et al.* Coral reefs in the Anthropocene. *Nature* **546**, 82–90 (2017).
25. Hughes, T. P. *et al.* Global warming and recurrent mass bleaching of corals. *Nature* **543**, 373–377 (2017).
26. Anthony, K. R. N. Coral reefs under climate change and ocean acidification: challenges and opportunities for management and policy. *Annu. Rev. Environ. Resour.* **41**, 59–81 (2016).
27. Cooley, S. R., Kite-Powell, H. L. & Doney, S. C. Ocean acidification's potential to alter global marine ecosystem services. *Oceanography (Wash. D.C.)* **22**, 172–181 (2009).
28. Pascal, N. *et al.* Economic valuation of coral reef ecosystem service of coastal protection: a pragmatic approach. *Ecosyst. Serv.* **21**, 72–80 (2016).
29. Pendleton, L. H. Valuing coral reef protection. *Ocean Coast. Manage.* **26**, 119–131 (1995).
30. Spalding, M. D. & Brown, B. E. Warm-water coral reefs and climate change. *Science* **350**, 769–771 (2015).

Supplementary Information is available in the online version of the paper.

Acknowledgements We thank R. Dunbar for the use of his laboratory and D. Mucciarone for laboratory assistance; the Australian Institute of Marine Science for scientific and technical support; and the following people for their support in the field and/or laboratory: M. Byrne, T. Hill, L. Caldeira, R. Johnson, D. Ross and the staff of the One Tree Island Research Station. Expedition and staff support was provided by the Carnegie Institution for Science. Additional support for staff, but not expedition expenses, was provided by the California Academy of Sciences and the Fund for Innovative Climate and Energy Research. This work was permitted by the Great Barrier Reef Marine Park Authority under permit G14/36863.1.

Author Contributions R.A., Y.T. and K.C. conceived and designed the project, conducted pilot studies, and collected preliminary data. R.A., Y.T., D.A.K., A.N., K.W., T.R., Y.N., J.Y. and K.C. performed the experiments. R.A. and K.C. performed computational analyses. R.A. wrote the manuscript with input from Y.T., D.A.K. and K.C. All co-authors reviewed and approved the final manuscript.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Publisher's note: Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations. Correspondence and requests for materials should be addressed to R.A. (ralbright@calacademy.org).

Reviewer Information *Nature* thanks H. Kayanne, J. Lough and the other anonymous reviewer(s) for their contribution to the peer review of this work.

METHODS

Experimental setup. Prior to low tide each day, a 15-m³ floating 'header' tank was partially submerged in the first lagoon, adjacent to the reef flat study site (Fig. 1). The tank was gravity-fed with ambient seawater from the lagoon. Two marine grade bilge pumps (3,000 gph, Five Oceans) were secured inside the tank for mixing during chemical addition and to deliver the solution to the study site. On 30 days, 4 g Rhodamine WT (20 g of a 20% solution, Turner Designs 10-108), dissolved in 0.5 l of reverse osmosis water, was manually added to the tank over the course of approximately 30 min and mixed. On 20 of those days, approximately 1 kg CO₂ was also introduced to the tank via a rosette of 12 airstones (Pentair AS30L), reducing the pH inside the tank from ambient (approximately 8.1) to approximately 6.1 (Extended Data Fig. 2). The solution inside the tank was subsequently mixed for an additional 30–45 min to ensure homogeneity. The tank was covered to minimize gas exchange. Tank pH was monitored by continuously pumping seawater from the tank through a flow cell containing a Honeywell Durafet pH sensor (Extended Data Fig. 2) and was stable to <0.03 throughout each experiment. The pH content of the header tank was intentionally low to account for dilution upon release to the reef flat. The relevant pH is the pH as the water enters our experimental zone. In the study site, the seawater pH was lowered by just 0.143 ± 0.007 units (mean \pm s.e.m.), from a background value of 8.126 ± 0.013 to a treatment level of 7.983 ± 0.015 .

The seawater solution from the tank (experiment days: seawater + CO₂ + dye; control days: seawater + dye) was pumped onto the reef flat at a constant rate (approximately 31 s^{-1}) for a period of 60 min starting at the predicted time of low tide (Extended Data Table 1). The solution was introduced to the study site via the two bilge pumps that were submerged in the tank, connected to two lengths of vinyl tubing (1.5-inch (3.8-cm) inner diameter) that were secured to a cinder block located approximately 15-m upstream from the centre of the study site. Throughout the addition, the ratio of CO₂ to dye being added to the study site was assumed to be constant, and flow within the study site was considered to be in steady state. With this study methodology, the part of the study site that is not affected by CO₂ enrichment also serves as control for the part of the study site that is affected by the plume.

The mean transit time of seawater across the study site was approximately 30 min. To ensure steady state, the seawater solution from the tank was pumped onto the reef flat for a period of approximately 60 min after peak low tide. After the 60-min pumping period, discrete water samples were taken at defined sampling locations along the length of two parallel transects that defined the borders of the study area, one along the upstream edge and the other along the downstream edge (Fig. 1). The upstream and downstream transects were separated by approximately 12.4 m. The length of each transect was 32 m. Thus, the study area consisted of a rectangle approximately 12.4 m by 32 m (approximately 400 m²). The upstream transect consisted of 11 sampling stations spanning the width of 32 m, and the downstream transect consisted of 15 sampling stations spanning the width of 32 m. Sample locations were strategically assigned with a higher density near the centre of the study area to accurately characterize the shape of the resulting alkalinity and dye plume/curve (Extended Data Fig. 1).

Discrete samples were collected at each of the 26 sample locations by pumping reef water into 500-ml borosilicate glass bottles (Corning, 1500-500 PYREX glass reagent bottle) using battery-operated liquid transfer pumps (Sierra Tools, Model JB5684). To minimize chemical variation owing to minor changes in sampling depth and/or location, precise sample locations were marked with plastic discs, nailed to the reef substrate. Samples were collected along the upstream and downstream transects simultaneously by five individuals, with each person sampling between five and six locations. All samples were typically collected in less than 3 min, and it was assumed that the study site was in steady state during this time (that is, all fluxes and flows did not change during the 3-min sampling interval). Samples were immediately returned to the One Tree Island Research Station, where they were subsampled and analysed for pH, total alkalinity (A_T), and Rhodamine WT (see 'Chemical analyses' section).

Five CTDs (YSI Models 6600, 6920) were deployed throughout the reef flat area, one upstream, two downstream and two outside the study area (Fig. 1). These instruments logged continuously at 2-min intervals over the thirty study days. A SeapHOx³¹ and SAMI p_{CO_2} sensor were deployed along the upstream edge of the study area for continuous measurements of seawater temperature, photosynthetically active radiation (PAR), salinity, depth, pH, p_{CO_2} and dissolved oxygen concentration to characterize natural variability in the system (Extended Data Fig. 7). These instruments logged at 10-min intervals over the 30 study days. NO₃⁻ and NH₃ concentrations were measured daily at a total of six stations, three upstream and three downstream. While A_T can be altered by the uptake and release of PO₄³⁻, NO₃⁻ and NH₄⁺, which are associated with photosynthesis, remineralization³² and by sulfate reduction^{33,34}, these fluxes should be of minor importance

in most reef settings^{35,36}. Salinity and nutrient data from our study are provided in Extended Data Table 2.

Alkalinity–dye slopes (r) and mean background alkalinities (\bar{a}) for each day were calculated using paired alkalinity and dye measurements that were collected across all sampling stations, transects and days. We conducted our experimental protocol 31 times, 21 times on experiment days and 10 times on control days. One experiment day (8 October 2016) was omitted from analyses because the upstream transect became undersaturated with respect to Ω_{arag} . We restricted analyses to days for which Ω_{arag} was supersaturated so that our estimated effect of depressed Ω_{arag} on NCC was not driven by thermodynamically controlled dissolution of calcium carbonate in undersaturated water. The remaining 30 days resulted in a total of 780 paired alkalinity and dye measurements that were used in the fitting procedure described below.

Hypotheses. We used the change in ratios between seawater alkalinity and a non-reactive dye (Rhodamine WT) to assess whether CO₂ enrichment reduces the NCC of a coral reef flat. By comparing the alkalinity–dye ratios before (that is, upstream of the study site), and after (that is, downstream) the water mass is influenced by the reef, we are able to estimate the change in alkalinity uptake (that is, NCC) that is associated with CO₂ enrichment, testing the hypotheses described below.

Null hypothesis, H₀: CO₂ enrichment does not affect NCC. If the reef does not respond to CO₂ enrichment, the change in alkalinity between the upstream and downstream transects would be due solely to background reef calcification. In this scenario, the change in alkalinity between the upstream and downstream transects would not be systematically related to the dye concentration; consequently, the ratio of the alkalinity–dye relationship (that is, the alkalinity–dye slope, r), would not be expected to change between the upstream and downstream locations (that is, $r_{\text{up}} = r_{\text{down}}$). The drawdown in alkalinity between the upstream and downstream locations is due to background reef calcification and results in a difference between the y-intercepts, \bar{a} (that is, $\bar{a}_{\text{up}} - \bar{a}_{\text{down}}$).

Alternate hypothesis, H₁: CO₂ enrichment reduces NCC. If CO₂ suppresses NCC, the drawdown in alkalinity is smaller in areas with more CO₂ (and more dye) than in areas with less CO₂ (and less dye). This effect yields a positive correlation between dye and alkalinity (that is, a positive alkalinity–dye slope) that increases as the water mass moves across the reef—in other words, the alkalinity–dye slope at the downstream transect is greater than that of the upstream transect ($r_{\text{down}} > r_{\text{up}}$).

Mathematical explanation: method for determining the alkalinity–dye relationships for each transect (upstream or downstream) by day. *Objective.* To test whether the CO₂ enrichment alters NCC, we want to test whether the alkalinity–dye slopes are significantly different between the upstream and downstream transects. Thus, the goal of the calculations described below is to obtain the best possible estimates of r by accounting for all known sources of variability as has been previously described¹⁰. In the following multivariate regression approach, known sources of error (including spatial and temporal variability) are accounted for to better constrain r . In this approach, r is calculated as a function of the measured alkalinity and dye, and station location and day are also included in the model as explanatory variables.

Multivariate regression. Nomenclature. 'Background' (for example, background alkalinity, background NCC, background Ω_{arag}) refers to levels that were (or would have been) present in the absence of CO₂ enrichment. The background values come from the intercepts of the regression lines where dye levels are 0.

Accounting for background spatial variability. To obtain more accurate estimates of r , background spatial variability in alkalinity can be taken into consideration. To account for this non-random variation as a component of the alkalinity signal, we devised a system in which each sampling location (s) was allocated a unique offset (x_s) corresponding to the departure of that particular location from the overall mean of the transect. Because calcification changed from day to day, the magnitude of these offsets also varied by day, by the value y_d . Thus, the offset for a particular station, on a given day, is given by the term $x_s y_d$. The mean magnitude of these offsets, provided in Extended Data Fig. 3, was small relative to the total alkalinity, but was included in subsequent analyses to obtain more accurate estimates of r (a more complete description of this method has previously been published¹⁰).

Method. The aim of this calculation is to determine: r_d , the alkalinity–dye slope for day d ; x_s , the average offset of background alkalinity at station s from the mean background alkalinity along the transect; and y_d , the amplitude of the pattern offsets on day d . With these values, and alkalinity and dye observations from a single station and day (that is, A_{sd} and D_{sd}), the mean background alkalinity along the transect for day d could be estimated using the equation

$$a_{sd} = A_{sd} + x_s y_d - r_d D_{sd}$$

The mean background alkalinity, \hat{a}_d , along the transect for day d can be estimated as the mean of the estimates based on each individual station for that day

$$\hat{a}_d = \text{mean}_s(a_{sd})$$

The mean of station offsets along a transect is zero

$$0 = \text{mean}_s(x_s)$$

The average amplitude of the pattern offset across all days equals 1

$$1 = \text{mean}_d(y_d)$$

If there were no additional unaccounted-for sources of variability, then a_{sd} would equal \hat{a}_d for all stations on day d . Therefore, x_s , y_d and r_d are determined by finding the values of these parameters that minimize the mean variance in $a_{sd} - \hat{a}_d$ across all stations and days, subject to the above constraints.

Calculated values for x_s , y_d , r_d and \hat{a}_d are shown in Extended Data Fig. 3. The central values shown in Extended Data Fig. 3c–h result from the ordinary least-squares curve fit to Supplementary Equation (1). The standard errors shown in Extended Data Fig. 3c–h are the square root of the sum of the squares of the standard errors resulting from the least-squares curve fit for Supplementary Equation (1) and the squares of the standard deviation of the central values obtained using a randomized choice of control days. In this process, we recalculated the central values 10,000 times using a bootstrapping method in which we chose ten control days with replacement randomly from among the ten actual control days.

Additional notes on calcification equations. Additional information for equations (1)–(3). Pumping rate of the dye, P_{dye} , was calculated

$$P_{\text{dye}} = D_{\text{tank}} \times P$$

in which D_{tank} is the dye concentration in the tank in g m^{-3} , and P is the pumping rate in $\text{m}^3 \text{s}^{-1}$. P was calculated daily by measuring the time (in seconds) needed to fill a known volume (120 l). The volumetric flow rate, F , was calculated as

$$F = P_{\text{dye}}/D_{\text{down}}$$

in which D_{down} is the mean dye concentration at the downstream transect in g m^{-3} .

Chemical analyses. Discrete samples were immediately returned to the laboratory on One Tree Island where they were analysed for pH_{total} , A_T and Rhodamine WT, and subsampled for the later determination of salinity and nutrients (NH_4^+ , NO_2^- and NO_3^-). All measurements and calculations were consistent with best practice recommendations³⁷. For 99.9% of station–day combinations (26 stations \times 30 days = 780 bottles), we successfully measured pH_{total} , A_T and Rhodamine WT, resulting in 779 paired measurements that were used in the fitting procedure described in the Supplementary Information.

Total dissolved inorganic carbon, Ω_{arag} , concentration of carbonate ions and p_{CO_2} were calculated as a function of A_T , pH_{total} , *in situ* salinity and temperature using previously published dissociation constants for carbonate^{38,39} and boric acid⁴⁰. All chemistry data are included in Supplementary Table 1. Nutrient concentrations were determined to be negligible in all samples (Extended Data Table 2).

A_T . Samples for A_T were pre-filtered using 0.45- μm filters (Whatman) and analysed in triplicate using a Metrohm 855 Robotic Titrator (Metrohm USA) using certified 0.1 M HCl (Fisher Chemical) diluted to a nominal concentration of 0.0125 M. Acid was calibrated daily by analysing Certified Reference Material (CRM, Batch 154) from the laboratory of A. Dickson. Instrumental drift was corrected by analysing an internal standard (made by filtering seawater to 0.45 μm) every tenth sample. A_T calculations were performed as previously described⁴¹. A_T by volume ($\mu\text{mol l}^{-1}$) was converted to A_T by mass ($\mu\text{mol kg}^{-1}$) by applying a density correction using salinity and temperature. The resulting mean and s.e.m. were calculated for each sample location on each day. Alkalinities were normalized to *in situ* salinities and salinity-normalized alkalinities were used for determination of alkalinity–dye relationships and NCC rates.

pH. Seawater pH from the discrete samples was analysed spectrophotometrically using *m*-cresol purple as the indicator dye using an automated system based on a previously published design⁴². An Ocean Optics Jazz spectrophotometer was used, and the temperature of the jacketed 10-cm cell was maintained at 25 °C using a water bath (Fisher Scientific, Model 4100 R20F). Duplicate bottles were collected at four locations every day to assess the reproducibility of the measurements, which was determined to be -0.0002 ± 0.0025 ($n = 91$). Purified *m*-cresol purple was used for most of the expedition to ensure accuracy; dye that we obtained directly from the manufacturer was corrected for impurities following previously published protocols⁴³. All pH values are reported on the total scale, and *in situ* pH was calculated using CO2SYS.

Rhodamine WT. Rhodamine WT concentration was measured fluorometrically using a Turner 10AU fluorometer and 25 ml cuvettes. A series of 8 standards were made by mass-diluting a 400 parts per 10⁹ (ppb) Rhodamine WT standard (Turner Designs) to 0, 0.5, 1, 2, 4, 16, 32 and 64 ppb. The standard curve was measured at the beginning and end of each measuring day to check for drift. Water samples were kept in a temperature-controlled water bath (Thermo Scientific, Precision Microprocessor Controlled 280 Series) at 25 °C before analysis to minimize temperature-induced errors in fluorescence. The temperature of each sample was recorded immediately after analysis using a digital thermometer accurate to ± 0.05 °C (VWR, Traceable Platinum Ultra-Accurate Digital Thermometer). Rhodamine WT concentrations were temperature-corrected using the formula $F_r = F_s e^{k(T_s - T_r)}$, here F_r and F_s are the fluorescence at the reference and sample temperatures, T_r and T_s , and $k = \frac{0.026}{K}$ equating to a 2.6% correction per K (ref. 44). Temperature corrections were applied before normalizing values to the standard curve. Dye concentrations were then normalized to salinity and salinity-normalized concentrations were used for subsequent analyses. Instrumental precision from triplicate measurements of this system was less than 0.1 ppb.

Nutrients. Nutrient samples were subsampled into 15-ml conical centrifuge tubes (Falcon). NH_3 samples were immediately frozen, and total ammonia concentrations (total $\text{NH}_3 = \text{NH}_3 + \text{NH}_4^+$) were later determined using a modified fluorometric method⁴⁵. NO_3^- samples were preserved with 0.1 ml 1 M HCl, closed, shaken and left in the dark at room temperature until transport to Eilat. NO_2^- was measured using a colorimetric method⁴⁶, using a Flow Injection Autoanalyzer (Lachat Instruments Model QuickChem 8500). NO_3^- was measured by reducing it to NO_2^- using a copperized cadmium column. Precision of NH_3 , NO_2^- and NO_3^- measurements was approximately $0.05 \mu\text{mol l}^{-1}$. NO_2^- and NO_3^- in this study are reported as total oxidized nitrogen ($\text{NO}_2^- + \text{NO}_3^-$). Results are provided in Extended Data Table 2.

Salinity. Discrete water samples were taken each day at each of the five CTD locations. Samples were stored in an air-conditioned, shaded room until transport to the Australian Institute of Marine Science for analysis on a Guildline Portasal Salinometer (Model 8410A), with precision of ± 0.0001 units. Accuracy was verified using Certified Reference Material (OSIL, IAPSO Standard Seawater). Results are provided in Extended Data Table 2.

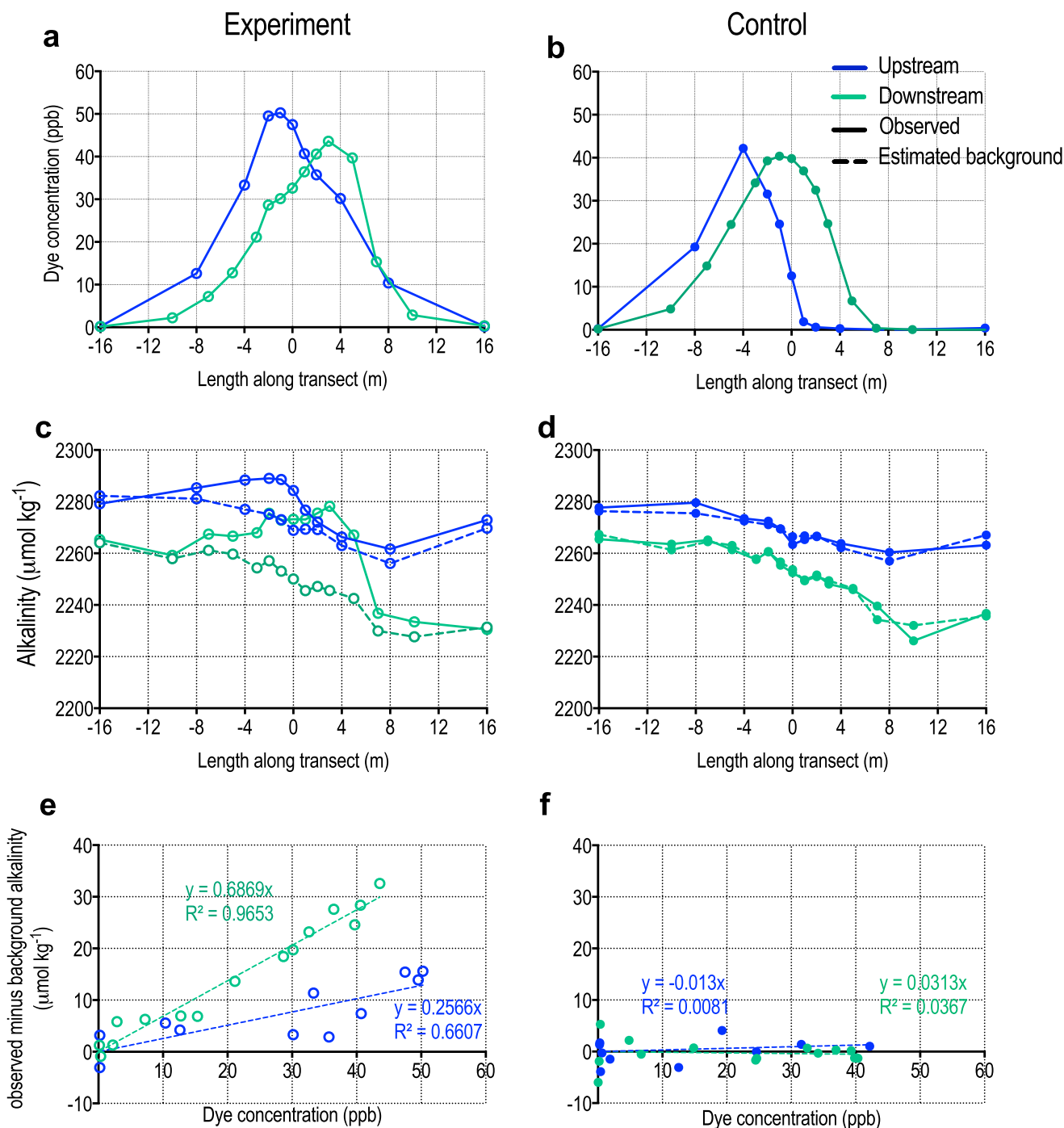
Benthic community structure. Benthic surveys were conducted to characterize the underlying community structure of the study area. Five, 12.4-m transects were laid on the reef flat perpendicular to the reef front, spaced approximately 8 m apart. Photographs were taken of 0.25-m² quadrats at 1-m intervals. Photographs were analysed using Coral Point Count software with Excel extensions (CPCe) using 15 random points per quadrat. The benthos was assigned to one of six categories: (1) live coral; (2) macroalgae; (3) turf algae; (4) crustose coralline algae (CCA); (5) sand/rubble; and (6) other. Where morphologic forms of CaCO_3 (for example rubble and CaCO_3 rock) were covered with biologically active groups (for example turf or CCA), the biologically active group was scored. Percentage cover by benthic type is as follows: crustose coralline algae (26%), live coral (15%), turf algae (28%), macroalgae (7%), sand/rubble (21%), and other (<3%).

Code availability. The Mathematica routine used to calculate the alkalinity–dye ratios (slopes) and dye-free mean alkalinity estimates (*y*-intercepts) for each day is provided in the Supplementary Information.

Data availability. The authors declare that all data supporting the findings of this study are available within the paper and its Supplementary Information.

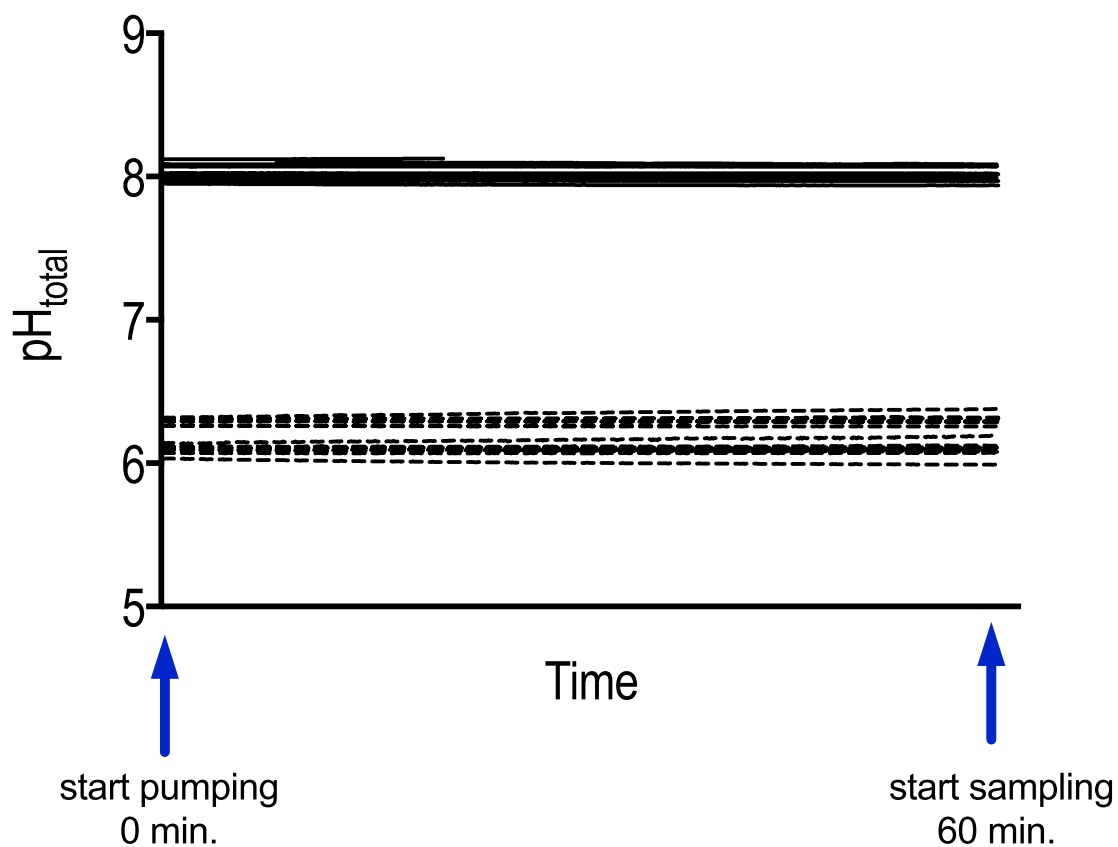
- Bresnahan, P. J., Martz, T. R., Takeshita, Y., Johnson, K. S. & LaShomb, M. Best practices for autonomous measurement of seawater pH with the Honeywell Durafet. *Methods Oceanogr.* **9**, 44–60 (2014).
- Brewer, P. G. & Goldman, J. C. Alkalinity changes generated by phytoplankton growth. *Limnol. Oceanogr.* **21**, 108–117 (1976).
- Berner, R. A., Scott, M. R. & Thomlinson, C. Carbonate alkalinity in the pore waters of anoxic marine sediments. *Limnol. Oceanogr.* **15**, 544–549 (1970).
- Gaines, A. G. & Pilson, M. E. Q. Anoxic water in the Pettaquamscutt River. *Limnol. Oceanogr.* **17**, 42–50 (1972).
- Kinsey, D. W. Alkalinity changes and coral reef calcification. *Limnol. Oceanogr.* **23**, 989–991 (1978).
- Gattuso, J. P., Allemand, D. & Frankignoulle, M. Photosynthesis and calcification at cellular, organismal and community levels in coral reefs: a review on interactions and control by carbonate chemistry. *Am. Zool.* **39**, 160–183 (1999).
- Dickson, A. G., Sabine, C. L. & Christian, J. R. *Guide to Best Practices for Ocean CO₂ Measurements* 191 (PICES Special Publication, 2007).
- Mehrbach, C., Culbertson, C. H., Hawley, J. E. & Pytkowicz, R. M. Measurement of the apparent dissociation constants of carbonic acid in seawater at atmospheric pressure. *Limnol. Oceanogr.* **18**, 897–907 (1973).
- Dickson, A. G. & Millero, F. J. A comparison of the equilibrium constants for the dissociation of carbonic acid in seawater media. *Deep-Sea Res. A* **34**, 1733–1743 (1987).
- Dickson, A. G. Thermodynamics of the dissociation of boric acid in synthetic seawater from 273.15 to 318.15 K. *Deep-Sea Res. A* **37**, 755–766 (1990).

41. Dickson, A. G., Afghan, J. D. & Anderson, G. C. Reference materials for oceanic CO₂ analysis: a method for the certification of total alkalinity. *Mar. Chem.* **80**, 185–197 (2003).
42. Carter, B. R., Radich, J. A., Doyle, H. L. & Dickson, A. G. An automated system for spectrophotometric seawater pH measurements. *Limnol. Oceanogr. Methods* **11**, 16–27 (2013).
43. Liu, X., Patsavas, M. C. & Byrne, R. H. Purification and characterization of meta-cresol purple for spectrophotometric seawater pH measurements. *Environ. Sci. Technol.* **45**, 4862–4868 (2011).
44. Wilson, J. F. in *Techniques for Water Resources Investigations of the U.S. Geological Survey, Book 3* (U.S. Government Printing Office, 1968).
45. Holmes, R. M., Aminot, A., Kérouel, R., Hooker, B. A. & Peterson, B. J. A simple and precise method for measuring ammonium in marine and freshwater ecosystems. *Can. J. Fish. Aquat. Sci.* **56**, 1801–1808 (1999).
46. Grasshoff, K., Kremling, K. & Ehrhardt, M. *Methods of Seawater Analysis* (Wiley-VCH, 1999).

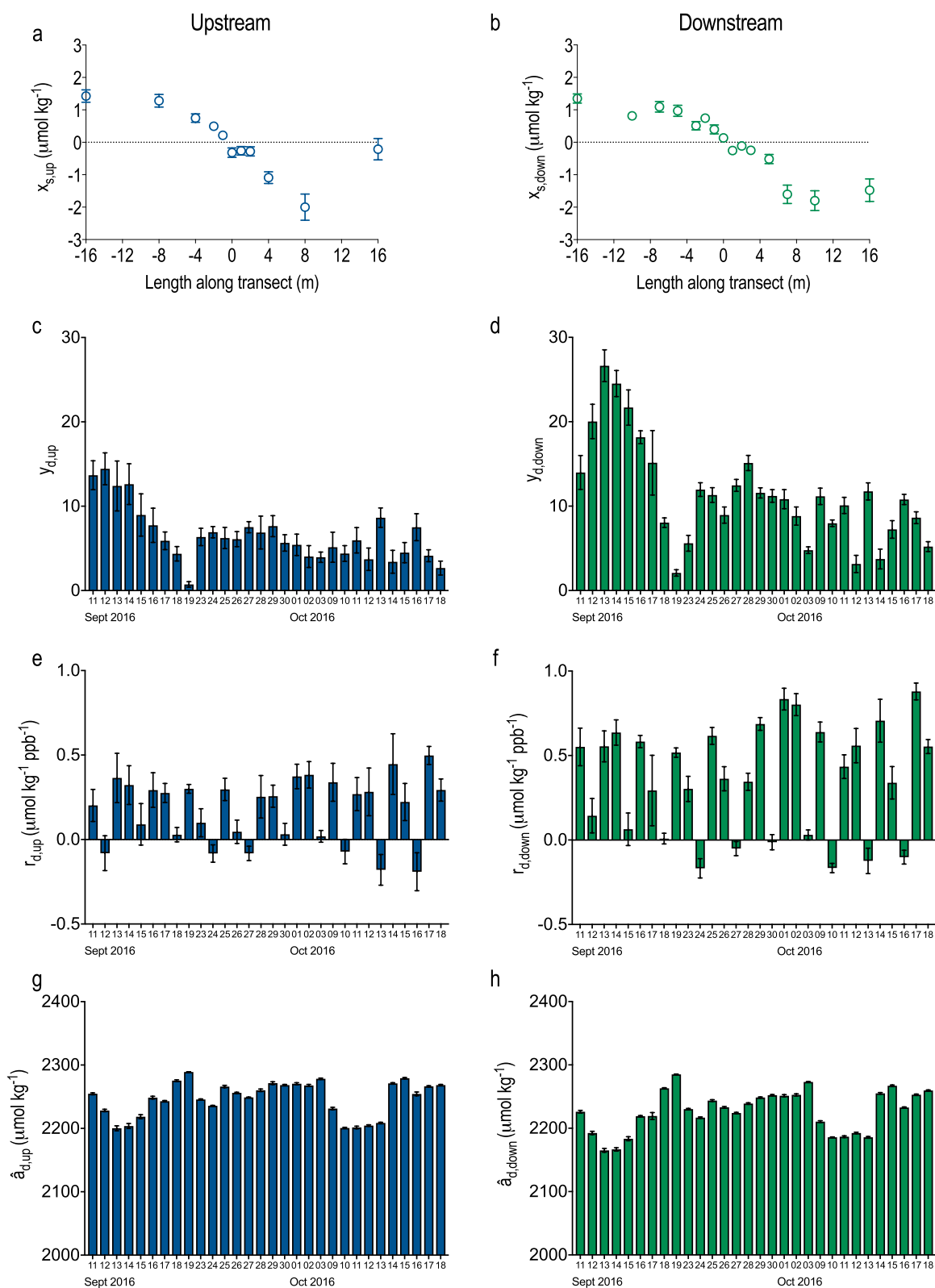


Extended Data Figure 1 | Relationships between alkalinity and dye for experiment and control days. a–f, Relationships between alkalinity and dye for a representative experiment day ($n = 20$ independent experiments, 29 September 2016 shown here) and control day ($n = 10$ independent experiments, 30 October 2016 shown here). a, b, Dye concentrations; c, d, alkalinities; e, f, alkalinity anomalies versus dye concentrations. Linear regressions were fit to alkalinity–dye data using least-squares residuals. d, On control days, the observed (measured) alkalinities closely agree with predicted values for each station. Comparing the upstream and downstream alkalinity–dye ratios provides an estimate of the effect of CO_2 enrichment on NCC, as described in the Methods. e, On experiment days

($n = 20$ independent experiments), if CO_2 suppresses NCC, the drawdown in alkalinity is smaller in areas with more CO_2 (and more dye) than in areas with less CO_2 (and less dye). This effect yields a positive correlation between dye and alkalinity (that is, a positive alkalinity–dye slope) that increases as the water mass moves across the reef—in other words, the alkalinity–dye slope at the downstream transect is greater than that of the upstream transect. f, On control days ($n = 10$ independent experiments), when dye but no CO_2 was added, alkalinity and dye were not correlated, and the mean alkalinity–dye slopes for the upstream and downstream transects did not differ from zero.

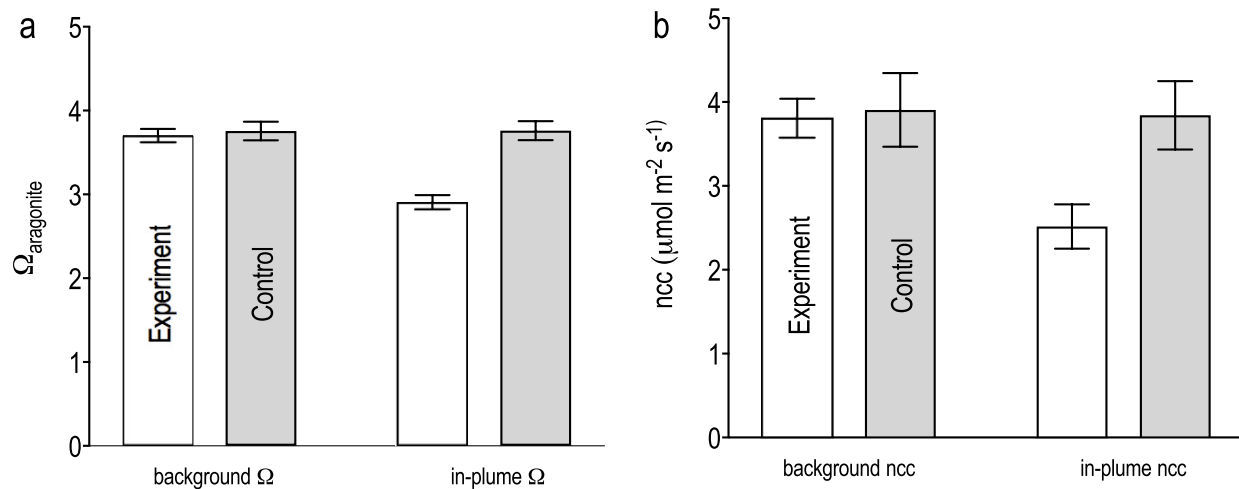


Extended Data Figure 2 | Time series of pH inside the header tank during the 60-min pumping period. Solid lines represent control days, and dashed lines represent experiment days.



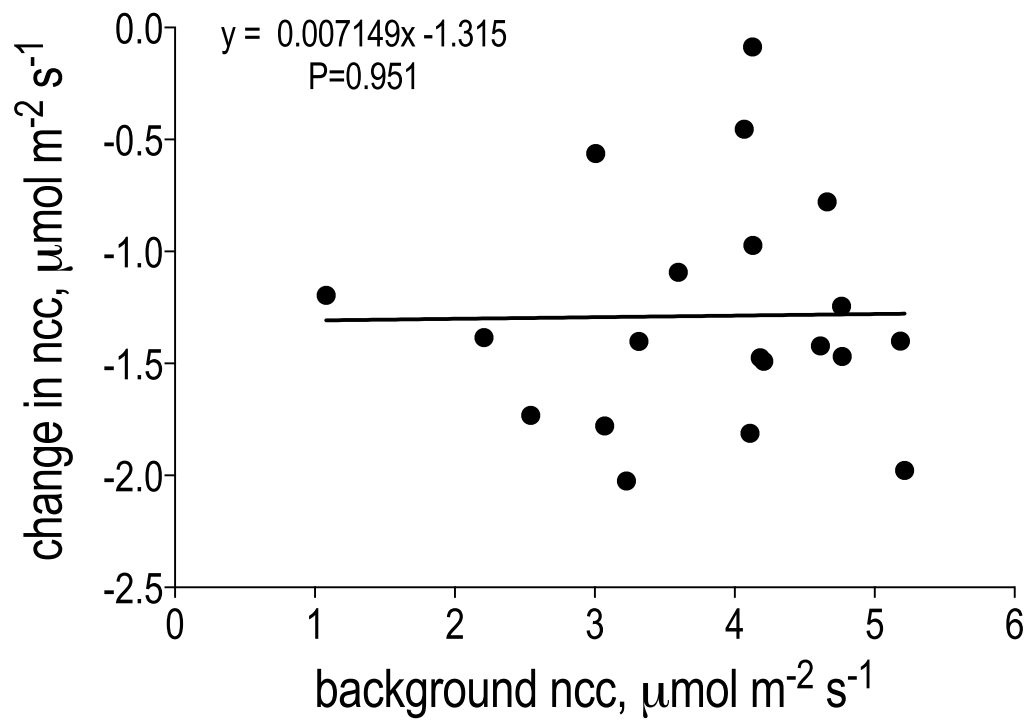
Extended Data Figure 3 | Results of the multivariate regression analysis. **a, b**, Unique offsets by station (x_s) for the upstream and downstream transects (mean \pm s.e.m., $n = 10$). **c, d**, Magnitude of offsets by day (y_d) for upstream and downstream transects. **e, f**, Alkalinity-dye ratios by day (r_d), for upstream and downstream transects. **g, h**, Mean background

alkalinities by day (\hat{a}_d) for upstream and downstream transects. In **c–h**, bars represent central values as calculated by the multivariate regression described in the ‘Mathematical explanation’ section of the Methods, and error bars represent s.e.m. (for **c, e** and **h**, $n = 11$; for **d, f** and **h**, $n = 15$).

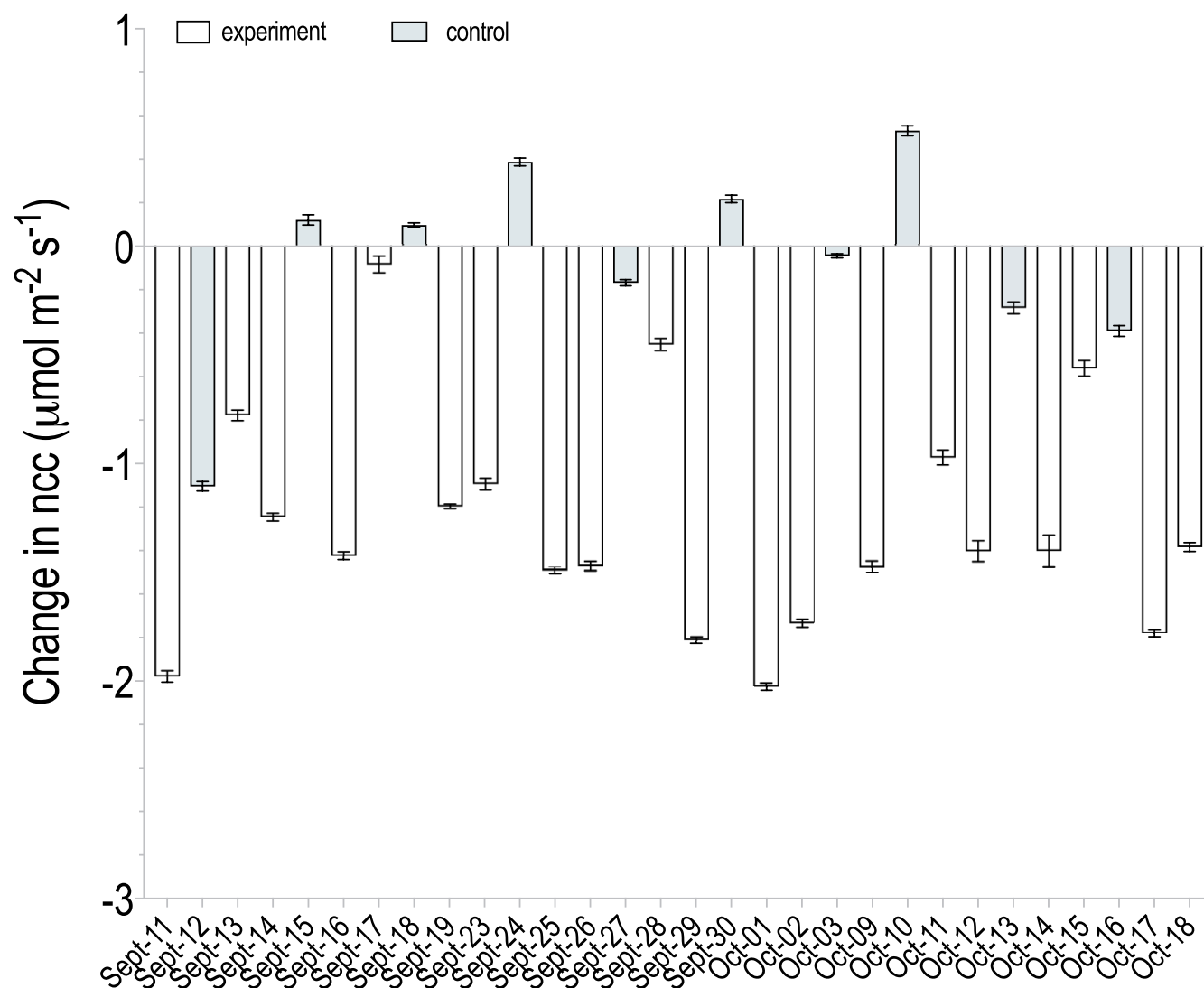


Extended Data Figure 4 | Mean Ω_{arag} and NCC rates for experiment and control days. Background and in-plume Ω_{arag} and NCC values (mean \pm s.e.m.) for experiment ($n = 20$ independent experiments)

and control ($n = 10$ independent experiments) days. Error bars reflect underlying natural variability (that is, day-to-day, hour-to-hour), because Ω_{arag} and NCC varied based on time of day and light availability.

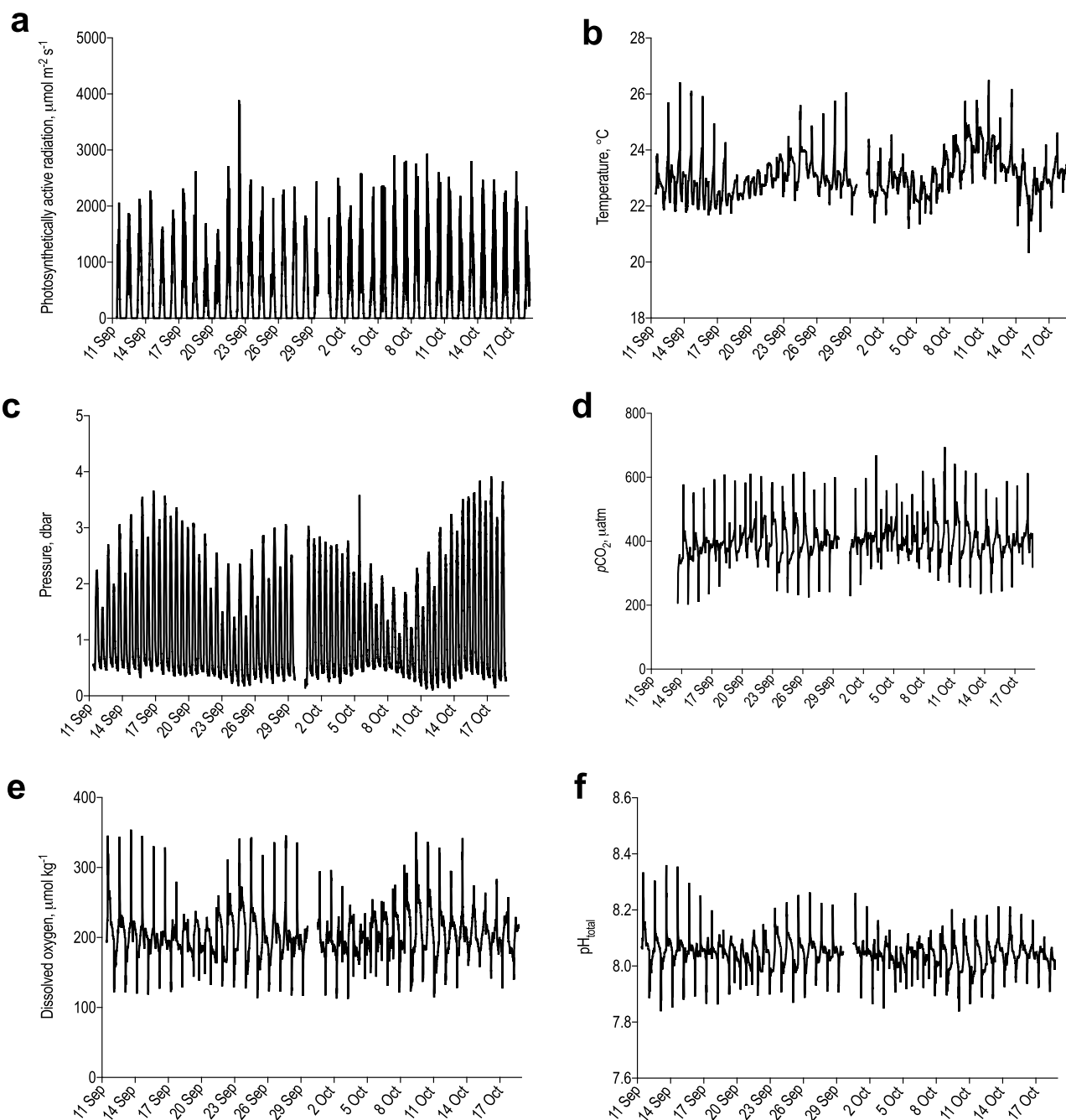


Extended Data Figure 5 | Relationship between the change in NCC and background NCC across all experiment days. Linear regression using least-squares residuals ($n = 20$ independent experiments).



Extended Data Figure 6 | Change in NCC by day. Change in NCC inside of the plume compared to background conditions (mean \pm s.e.m.) for experiment ($n=20$ independent experiments) and control ($n=10$ independent experiments) days. On all 20 experiment days, we detected a

statistically significant reduction in calcification within the plume (paired t -tests). On control days, in-plume NCC was higher than background NCC on five days, and lower than background NCC on five days.



Extended Data Figure 7 | Physical and chemical conditions of the study site. Time series of environmental data from SeapHOx and SAMI $p\text{CO}_2$ sensors. Instruments logged at 10-min intervals over the duration of the

study. Gaps in the data correspond to when the instruments were removed from the reef for maintenance. **a**, PAR; **b**, temperature; **c**, pressure; **d**, salinity; **e**, dissolved oxygen; **f**, pH.

Extended Data Table 1 | Dates, times, predicted heights of low tide and mean PAR for all days

| Date | Low Tide (HHMM) | Water Depth (m) | PAR ($\mu\text{mol m}^{-2} \text{s}^{-1}$) | Study Type |
|---------|-----------------|-----------------|----------------------------------------------|------------|
| 11 Sept | 1002 | 0.86 | 645 | Experiment |
| 12 Sept | 1104 | 0.72 | 1068 | Control |
| 13 Sept | 1149 | 0.53 | 1674 | Experiment |
| 14 Sept | 1236 | 0.36 | 1584 | Experiment |
| 15 Sept | 1317 | 0.20 | 1276 | Control |
| 16 Sept | 1356 | 0.10 | 1087 | Experiment |
| 17 Sept | 1436 | 0.07 | 535 | Experiment |
| 18 Sept | 1517 | 0.12 | 191 | Control |
| 19 Sept | 1601 | 0.24 | 89 | Experiment |
| 23 Sept | 0657 | 0.75 | 833 | Experiment |
| 24 Sept | 0822 | 0.86 | 1439 | Control |
| 25 Sept | 0949 | 0.82 | 466 | Experiment |
| 26 Sept | 1100 | 0.69 | 1968 | Experiment |
| 27 Sept | 1157 | 0.52 | 1634 | Control |
| 28 Sept | 1245 | 0.38 | 1421 | Experiment |
| 29 Sept | 1327 | 0.28 | 100 | Experiment |
| 30 Sept | 1406 | 0.24 | 1128 | Control |
| 01 Oct | 1443 | 0.27 | 849 | Experiment |
| 02 Oct | 1517 | 0.34 | 578 | Experiment |
| 03 Oct | 1551 | 0.46 | 169 | Control |
| 09 Oct | 0738 | 0.95 | 1251 | Experiment |
| 10 Oct | 0915 | 0.94 | 1708 | Control |
| 11 Oct | 1030 | 0.81 | 1244 | Experiment |
| 12 Oct | 1126 | 0.63 | 927 | Experiment |
| 13 Oct | 1214 | 0.44 | 1583 | Control |
| 14 Oct | 1258 | 0.28 | 1521 | Experiment |
| 15 Oct | 1341 | 0.18 | 677 | Experiment |
| 16 Oct | 1424 | 0.14 | 921 | Control |
| 17 Oct | 1507 | 0.18 | 732 | Experiment |
| 18 Oct | 1553 | 0.29 | 458 | Experiment |

Tide data were provided courtesy of One Tree Research Station. Low tide represents the time at which pumping onto the reef started; sampling occurred 60 min afterward. There was no significant difference between mean PAR (mean \pm s.e.m.) for treatment ($902 \pm 115 \mu\text{mol m}^{-2} \text{s}^{-1}$) and control days ($1,112 \pm 175 \mu\text{mol m}^{-2} \text{s}^{-1}$) days (unpaired *t*-tests).

Extended Data Table 2 | Mean \pm s.e.m. values for salinity, NH_4^+ and $\text{NO}_2^- + \text{NO}_3^-$ during and control days

| | Experiment (N=20) | | Control (N=10) | |
|--------------------------------------------------------|-------------------|------------------|------------------|------------------|
| | Upstream | Downstream | Upstream | Downstream |
| Salinity | 35.83 ± 0.02 | 35.86 ± 0.02 | 35.83 ± 0.02 | 35.84 ± 0.03 |
| NH_4 ($\mu\text{mol l}^{-1}$) | 0.37 ± 0.02 | 0.31 ± 0.02 | 0.38 ± 0.02 | 0.30 ± 0.02 |
| $\text{NO}_2 + \text{NO}_3$ ($\mu\text{mol l}^{-1}$) | 1.25 ± 0.09 | 1.08 ± 0.08 | 1.23 ± 0.09 | 1.16 ± 0.10 |

Note that underlying natural variability (that is, day-to-day, hour-to-hour) contributes to s.e.m. Measurement errors for each parameter are indicated in the Methods. Experiment days, $n = 20$ independent experiments; control days $n = 10$ independent experiments.

Dysregulation of expression correlates with rare-allele burden and fitness loss in maize

Karl A. G. Kremling¹, Shu-Yun Chen^{2,3}, Mei-Hsiu Su², Nicholas K. Lepak⁴, M. Cinta Romay², Kelly L. Swarts^{1,5}, Fei Lu^{2,6}, Anne Lorant⁷, Peter J. Bradbury⁴ & Edward S. Buckler^{1,2,4}

Here we report a multi-tissue gene expression resource that represents the genotypic and phenotypic diversity of modern inbred maize, and includes transcriptomes in an average of 255 lines in seven tissues. We mapped expression quantitative trait loci and characterized the contribution of rare genetic variants to extremes in gene expression. Some of the new mutations that arise in the maize genome can be deleterious; although selection acts to keep deleterious variants rare, their complete removal is impeded by genetic linkage to favourable loci and by finite population size^{1–4}. Modern maize breeders have systematically reduced the effects of this constant mutational pressure through artificial selection and self-fertilization, which have exposed rare recessive variants in elite inbred lines⁵. However, the ongoing effect of these rare alleles on modern inbred maize is unknown. By analysing this gene expression resource and exploiting the extreme diversity and rapid linkage disequilibrium decay of maize⁶, we characterize the effect of rare alleles and evolutionary history on the regulation of expression. Rare alleles are associated with the dysregulation of expression, and we correlate this dysregulation to seed-weight fitness. We find enrichment of ancestral rare variants among expression quantitative trait loci mapped in modern inbred lines, which suggests that historic bottlenecks have shaped regulation. Our results suggest that one path for further genetic improvement in agricultural species lies in purging the rare deleterious variants that have been associated with crop fitness.

The phenotypic consequences of rare deleterious alleles are of interest for their role in disease and fitness, but their effects are difficult to detect without prohibitively large sample sizes^{4,7,8}. Gene expression, a phenotype for which millions of observations are obtainable, has previously been associated with rare alleles through pedigrees⁹ and correlated with the burden of rare alleles in putative *cis*-regulatory regions¹⁰. However, links between regulation of expression, fitness and altered allele frequencies resulting from population bottlenecks are not well established.

Maize provides a powerful system with which to evaluate these questions. Although the mutation rate of maize ($9\text{--}20 \times 10^{-9}$ mutations per base pair per generation) is similar to that of humans¹¹, the rapid decay of linkage disequilibrium in much of the genome is an order of magnitude faster than in other large eukaryotic genomes, which improves the resolution of associating phenotypes with both rare and common genotypes (average $r^2 < 0.1$ in 2 kb)^{6,12}. Maize is also very diverse (nucleotide diversity, $\pi \approx 1.4\%$ ¹³; approximately $14\times$ more than humans) having preserved much of the diversity from its wild tropical relative teosinte¹². However, population bottlenecks during temperate adaptation and modern breeding have severely reduced the diversity of maize and severely reduced its effective population size (N_e) from more than one million in pre-bottleneck tropical landraces^{5,14,15}.

Because it has undergone intensive selection and systematic inbreeding since approximately 1900, which has contributed to an eightfold increase in productivity^{16,17}, maize also provides an extreme case in which to examine the capacity of selection and inbreeding to purge deleterious variants. Additionally, the rapid linkage disequilibrium decay in maize enables the testing of whether rare alleles or novel combinations of common alleles drive extreme expression.

Here we quantified mRNA expression from 299 maize lines that represent the genotypic and phenotypic diversity of modern inbred maize¹⁸. We automated a 3' mRNA sequencing method (QuantSeq, Lexogen GmbH), which is more efficient and accurate than mRNA sequencing and deals well with paralogues^{19,20}. We used this method to profile seven diverse tissues (Extended Data Fig. 1 and Supplementary Table 1). All lines had previously been genotyped by whole genome sequencing to produce a set of 61,430,377 segregating variants²¹, which we used to map expression quantitative trait loci (eQTL). We also calculated the fraction of variance in expression explained by individual single nucleotide polymorphisms (SNPs). We further used the variance explained to quantify whether alleles that were rare in pre-bottleneck tropical lines²² have a disproportionately large role in the regulation of expression in modern temperate bottlenecked germplasm (see Methods).

To determine whether rare alleles contribute to extreme gene expression, we investigated whether rare-allele abundance is greatest in individuals with extreme gene expression. We calculated an expression rank and upstream (5 kb) rare-allele count for each combination of gene and individual maize line¹⁰ (Fig. 1a). In each tissue, this was done separately for each of the 5,000 most highly expressed genes. At each expression rank we plotted the mean number of rare alleles across the 5,000 gene-line combinations, which gave a mean number of rare *cis* variants at each rank. If rare alleles drive extreme expression, then individuals at the lowest and highest expression ranks should have the highest number of rare *cis* variants. Significance was tested using a quadratic regression to compare the expression rank with the *cis* rare-allele count as previously described¹⁰. Using allele frequencies from previous whole genome sequencing of these lines²¹, we find that the abundance of rare SNPs within 5-kb upstream of the nearest gene (minor allele frequency (MAF) ≤ 0.05) is significantly correlated with extreme over- and underexpression in all tissues, relative to the population mean per gene ($P < 1.90 \times 10^{-80}$ and $R^2 = 0.72$, Fig. 1b and Extended Data Fig. 2). Although the quadratic regression is highly significant, even more-extreme departures exist at the tails than predicted by the regression (Fig. 1b, c); we therefore focus on the rare-allele count deviations of the five highest- and lowest-expressing individuals across the 5,000 most-expressed genes. Across the tissues, individuals that express at the lowest five expression ranks are enriched 1.68–1.86-fold for local rare alleles compared to the middle two quartiles, whereas the top five

¹Section of Plant Breeding and Genetics, 175 Biotechnology Building, Cornell University, Ithaca, New York 14853, USA. ²Institute for Genomic Diversity, 175 Biotechnology Building, Cornell University, Ithaca, New York 14853, USA. ³Institute of Plant and Microbial Biology, Academia Sinica 128, Sec 2nd, Academia road, Taipei, 11529, Taiwan. ⁴USDA-ARS, R. W. Holley Center, Cornell University, Ithaca, New York 14853, USA. ⁵Research Group for Ancient Genomics and Evolution, Department of Molecular Biology, Max Planck Institute for Developmental Biology, Spemannstr. 35, 72076 Tübingen, Germany. ⁶The State Key Laboratory of Plant Cell and Chromosome Engineering, Institute of Genetics and Developmental Biology, Chinese Academy of Sciences, Beijing 100101, China. ⁷Department of Plant Sciences, University of California Davis, Davis, California 95616, USA.

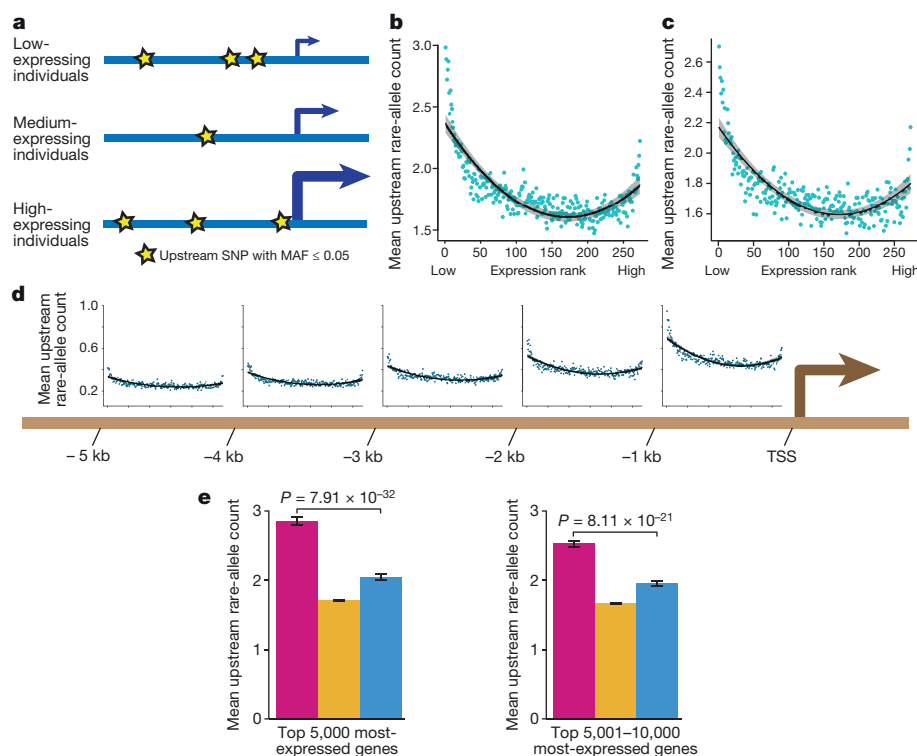


Figure 1 | The abundance of local rare alleles correlates with extremes in expression. **a**, A greater number of rare alleles is expected upstream of a gene in individuals that over- or underexpress a given gene relative to the mean of the population. **b**, Significant quadratic relationship between the expression rank of each line, in each of the top 5,000 most-expressed genes and the average local (5-kb upstream) rare-allele count. Quadratic regression¹⁰ of root tissue is shown here. ($n = 273$ unique inbred samples, $P = 1.95 \times 10^{-75}$, $R^2 = 0.72$). Each point is the mean number of rare alleles within 5-kb upstream of genes for lines that share an expression rank for one of the 5,000 genes. **c**, Quadratic relationship between expression rank for medium-expressed genes (5,001–10,000) and local (5-kb upstream) rare-allele count ($n = 273$ unique inbred samples, $P = 2.16 \times 10^{-66}$, $R^2 = 0.67$).

d, Expression ranks versus 5-kb upstream rare-allele counts divided into 1-kb windows ($n = 273$ unique inbred samples). TSS, transcription start site. **e**, Comparison of the number of rare *cis* alleles near genes for individuals in the bottom five expression ranks (fuchsia, $n = 5$ unique inbred samples measured for each of 5,000 genes) versus the middle two quartiles (yellow, $n = 137$ unique inbred samples measured for each of 5,000 genes) versus the top five expression ranks (blue, $n = 5$ unique inbred samples measured for each of 5,000 genes) (mean \pm s.e.m.) within the top 5,000 and next 5,000 most-expressed genes. P values from two-sided t -tests, comparing the top and bottom five expression ranks in the top 5,000 and next 5,000 most-expressed gene sets, are shown. This is consistent across tissues and gene sets (Extended Data Fig. 2, 3).

ranks are enriched by 1.22–1.46-fold over the middle two quartiles (Fig. 1b, e and Extended Data Fig. 2).

When broken down by proximity to the transcription start site, we observe that rare variants nearest to the transcription start site have the greatest effect (Fig. 1d). It is also notable that the ratio of underexpressing to overexpressing individuals possessing rare *cis* variants is greater nearest to the transcription start site (Fig. 1d). This is consistent with the proposition that disruption to gene-proximal promoters lowers expression, whereas distal disruptions have a relatively greater chance of increasing expression.

Although both over- and underexpression are potentially deleterious consequences of dysregulation, we note that there are significantly more *cis* rare alleles in underexpressing individuals than there are near the same genes in overexpressing individuals ($P = 7.91 \times 10^{-32}$, Fig. 1e). This observation is consistent with stronger selection against underexpression, and with the expression dosage model of fitness and heterosis²³. Additionally, it may be the case that rare variants are more likely to disrupt a promoter element than a repressor element, which would lead to a loss rather than a gain of expression. However, the greater apparent abundance of rare *cis* variants in underexpressing individuals may also partly be the result of under-calling expression, owing to mapping bias for genes that are located on rare haplotypes.

To evaluate how regulatory load affects genes at different expression levels, we compared how this load affects more- and less-expressed genes in each tissue. Purifying selection acts on regulatory variants²⁴ and is expected to be strongest near the most highly expressed genes²⁵. This can result in the purging of dysregulatory variants or—if selection

is not strong enough—an increase in rare variants because deleterious variants are reduced in frequency but not purged. By comparing rare-allele burdens 5-kb upstream of the most highly expressed genes (top 5,000) with those upstream of the medium-expressed genes (next 5,000 most-expressed genes), we note that across all tissues the 5,000 most highly expressed genes contain more *cis* rare alleles (Fig. 1e and Extended Data Figs 2, 3). However, consistent with an increased effect of purifying selection on the most highly expressed genes, we also note that the deleterious *cis* variants appear to be purged from the top 1,000 most highly expressed genes, which in some tissues have fewer rare *cis* variants than the next 1,000 most-expressed genes (Extended Data Fig. 4).

These analyses reveal the cumulative effects of rare alleles; we can also directly estimate the effect of ancestrally rare alleles on expression by exploiting the evolutionary history and bottlenecks in maize. This is possible because we can sample allele frequencies in extant tropical maize populations that have not experienced the same bottlenecks, which enables us to infer the ancestral allele frequency for each SNP. Approximately half a million SNPs in the maize HapMap3 (about 1% of total SNPs) that are rare (≤ 0.05 MAF) in 335 genotyped tropical lines²² are common (> 0.2 MAF) across our predominantly temperate expression-profiled lines (the base of the third leaf is shown in Fig. 2; remaining tissues shown in Extended Data Figs 5, 6), illustrating the effects of the tropical-to-temperate bottleneck. This class of formerly rare tropical SNPs—which is now of sufficient frequency to be easily detected in our eQTL association study using modern inbred lines—enables powerful tests of ancestrally rare-allele effects. By comparing

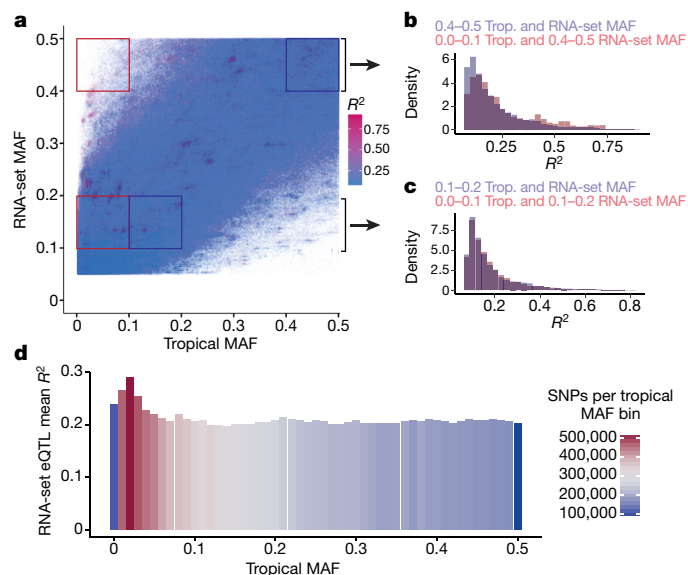


Figure 2 | Ancestral rare alleles are significantly enriched for highly explanatory *cis* eQTL in modern germplasm. **a**, Two-dimensional MAF plot coloured by R^2 from *cis* eQTL for SNPs associated below $P = 0.00001$. Using $n = 263$ unique inbred lines and conducting an eQTL association study for each expression trait, we quantified the fraction of variance in expression explained (R^2) by each *cis* SNP for its most strongly associated gene. eQTL R^2 was determined by linear regression (matrix eQTL). **b**, **c**, Across temperate MAF levels, significantly more variance in expression is explained by SNPs that are rare in the tropical (Trop.) maize germplasm that has not undergone a bottleneck (two-sided Wilcoxon signed-rank and Kolmogorov–Smirnov tests, $P < 1.12 \times 10^{-240}$). Results of tissue from base of leaf three shown here ($n = 263$ unique inbred samples). **d**, Histogram comparing RNA-set *cis* eQTL R^2 across tropical MAF bins for SNPs with MAF over 0.2 in the RNA set. R^2 determined by eQTL regression in matrix eQTL. Bars are coloured by the number of SNPs in each tropical MAF bin to illustrate that rarer alleles, which explain the most variance in expression, constitute the most abundant class of SNPs. A minimum RNA-set MAF of 0.2 is used, given the limited statistical power to map eQTL beneath this threshold with our RNA-set sample sizes. $n = 263$ unique inbred samples used in eQTL mapping.

the variance explained by alleles that are common in tropical maize lines versus those that are rare in these lines, while controlling for equal allele frequency in the RNA set, it is clear that tropical rare alleles explain significantly more variance in expression than do tropical common alleles across the whole genome in all tissues (Fig. 2b, c and Extended Data Figs 5, 6, Kolmogorov–Smirnov and Wilcoxon signed-rank tests, $P < 9.06 \times 10^{-33}$ for all tissues). This is clear from the high- R^2 SNPs, which can be seen to the left of the diagonal in the 2D MAF plot in Fig. 2a. Although some of these differences are adaptive, there are only a few hundred loci in the maize genome with a highly significant F_{st} (fixation index) that are thought to be adaptive^{12,26}. Therefore, many of the highly explanatory eQTL SNPs are likely to be deleterious. This highlights the fact that formerly rare alleles have significantly larger effects on quantitative expression phenotypes than do common variants, and suggests that the temperate bottleneck imposed substantial changes on gene regulation.

Finally, we investigated the association between dysregulation in gene expression and a fitness trait. To quantify fitness, we used previously published seed yields of the expression-profiled lines (multi-environment best linear unbiased predictions²⁷). First, we tested whether overall expression of the top 5,000 most-expressed genes could predict seed yields using ridge regression, which works across tissues (Extended Data Fig. 7). However, if dysregulation has a major role in altering phenotype, then the deviation in expression should also be predictive. We find support for this in two analyses: the cumulative deviation in expression for the 5,000 most-expressed genes in adult leaves and kernels significantly correlates with seed-weight fitness (Extended Data Fig. 8, Extended Data Table 1 and Supplementary Methods). Furthermore, by using the matrix of absolute deviation for the top 5,000 most-expressed genes in each tissue we were able to predict seed-weight fitness using ridge regression in six of seven tissues (Fig. 3). This predictive ability is significant in all tissues (Fig. 3, $r = 0.19–0.38$, $P = 0.01$ to $P < 2.9 \times 10^{-7}$) except for immature roots ($P = 0.68$). Consistent with expectations, an increased rare-allele burden does correlate with decreased fitness (Extended Data Fig. 9), but this result is likely to be partially confounded with the recent breeding history of maize and pedigree relationships. Large bi-parental populations without population structure would be better suited to addressing this question.

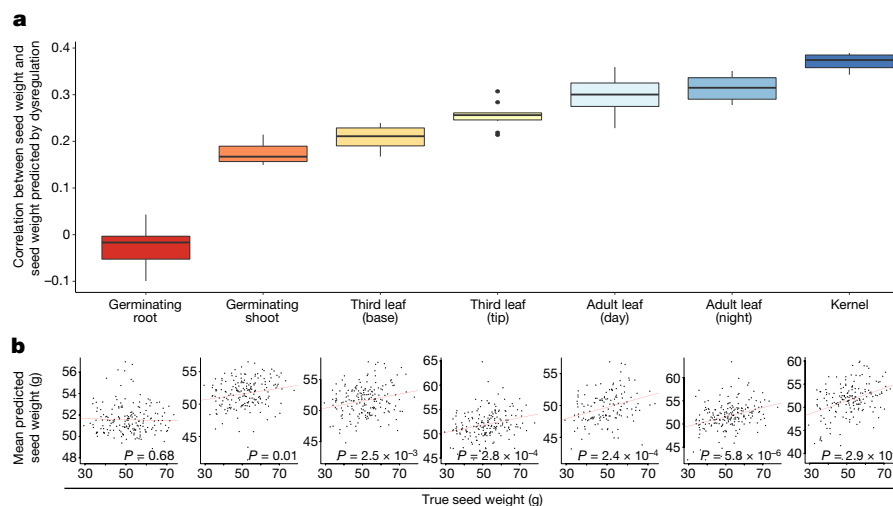


Figure 3 | Dysregulation of expression can predict fitness. Dysregulation of expression in the top 5,000 most-expressed genes from six of seven tissues significantly predicts seed-weight fitness. **a**, Range of correlations between predicted and true seed weight from ten repetitions of nested tenfold cross validation (ten inner and ten outer) using ridge regression. In the box plots, the middle horizontal lines represent the median, hinges represent the 25th and 75th percentiles (the interquartile range), the upper and lower whiskers extend to maximum and minimum points no more than $1.5 \times$ the interquartile range beyond the hinges, and individual

dots are outliers beyond the whiskers. **b**, True seed weight versus mean of predicted seed weight in grams. P values obtained from linear regression between true seed weight and mean of predicted seed weight. For both **a** and **b**, sample sizes were as follows: 2-cm root tips (unique $n = 181$) and shoots (unique $n = 183$) of germinating seedlings; 2-cm base (unique $n = 181$) and tip (unique $n = 182$) of leaf 3, leaves collected in the field during the day (unique $n = 135$) and night (unique $n = 187$); and 350-growing-degree-day kernels (unique $n = 171$), post sexual maturity (anthesis).

The ability of dysregulation to predict fitness is notable, given that the expression profiles were not collected in the same environments or years in which the seed-weight fitness phenotypes were determined. In brief, environmental variance cannot be controlled between the individuals in which expression and fitness were measured, which adds noise to these regressions. For this reason, the effect of expression dysregulation on fitness calculated here may be underestimated.

Our results demonstrate the influence of rare genetic variation on gene expression, and provide an example that connects the dysregulation of expression and a fitness trait. Consistent with population genetic expectations and evidence that recombination is insufficient to purge deleterious variants in modern maize²⁸, these results illustrate the disproportionate effect of rare alleles on thousands of expression phenotypes as well as the consequences of newly common alleles in modern low- N_e populations. Despite intensive selection and millions of yearly field trials by breeders⁵, maize provides evidence for the persistence of rare deleterious alleles in modern agricultural species after a strong bottleneck. This suggests that even intensive artificial selection is insufficient to purge genetic load. Although genomic selection has successfully combined favourable sets of common variants to improve yields, additional targeted breeding approaches and genetic manipulation would enable further removal of deleterious mutations and their phenotypic consequences.

Online Content Methods, along with any additional Extended Data display items and Source Data, are available in the online version of the paper; references unique to these sections appear only in the online paper.

Received 7 October 2016; accepted 1 February 2018.

Published online 14 March 2018.

- Kimura, M., Maruyama, T. & Crow, J. F. The mutation load in small populations. *Genetics* **48**, 1303–1312 (1963).
- Marth, G. T. et al. The functional spectrum of low-frequency coding variation. *Genome Biol.* **12**, R84 (2011).
- Henn, B. M., Botigüé, L. R., Bustamante, C. D., Clark, A. G. & Gravel, S. Estimating the mutation load in human genomes. *Nat. Rev. Genet.* **16**, 333–343 (2015).
- Gibson, G. Rare and common variants: twenty arguments. *Nat. Rev. Genet.* **13**, 135–145 (2012).
- Troyer, A. F. A retrospective view of corn genetic resources. *J. Hered.* **81**, 17–24 (1990).
- Remington, D. L. et al. Structure of linkage disequilibrium and phenotypic associations in the maize genome. *Proc. Natl Acad. Sci. USA* **98**, 11479–11484 (2001).
- Kono, T. J. Y. et al. The role of deleterious substitutions in crop genomes. *Mol. Biol. Evol.* **33**, 2307–2317 (2016).
- Manolio, T. A. et al. Finding the missing heritability of complex diseases. *Nature* **461**, 747–753 (2009).
- Li, X. et al. Transcriptome sequencing of a large human family identifies the impact of rare noncoding variants. *Am. J. Hum. Genet.* **95**, 245–256 (2014).
- Zhao, J. et al. A burden of rare variants associated with extremes of gene expression in human peripheral blood. *Am. J. Hum. Genet.* **98**, 299–309 (2016).
- Jiao, Y. et al. Genome-wide genetic changes during modern breeding of maize. *Nat. Genet.* **44**, 812–815 (2012).
- Gore, M. A. et al. A first-generation haplotype map of maize. *Science* **326**, 1115–1117 (2009).
- Tenaillon, M. I. et al. Patterns of DNA sequence polymorphism along chromosome 1 of maize (*Zea mays* ssp. *mays* L.). *Proc. Natl Acad. Sci. USA* **98**, 9161–9166 (2001).
- Vigouroux, Y. et al. Rate and pattern of mutation at microsatellite loci in maize. *Mol. Biol. Evol.* **19**, 1251–1260 (2002).
- Beissinger, T. M. et al. Recent demography drives changes in linked selection across the maize genome. *Nat. Plants* **2**, 16084 (2016).
- Duvick, D. N. The contribution of breeding to yield advances in maize (*Zea mays* L.). *Adv. Agron.* **86**, 83–145 (2005).
- Troyer, A. F. & Wellin, E. J. Heterosis decreasing in hybrids: yield test inbreds. *Crop Sci.* **49**, 1969–1976 (2009).
- Flint-Garcia, S. A. et al. Maize association population: a high-resolution platform for quantitative trait locus dissection. *Plant J.* **44**, 1054–1064 (2005).
- Eveland, A. L., McCarty, D. R. & Koch, K. E. Transcript profiling by 3'-untranslated region sequencing resolves expression of gene families. *Plant Physiol.* **146**, 32–44 (2008).
- Lohman, B. K., Weber, J. N. & Bolnick, D. I. Evaluation of TagSeq, a reliable low-cost alternative for RNAseq. *Mol. Ecol. Resour.* **16**, 1315–1321 (2016).
- Bukowski, R. et al. Construction of the third generation *Zea mays* haplotype map. *Gigascience* <http://doi.org/10.1093/gigascience/gix134> (2017).
- Romay, M. C. et al. Comprehensive genotyping of the USA national maize inbred seed bank. *Genome Biol.* **14**, R55 (2013).
- Yao, H., Dogra Gray, A., Auger, D. L. & Birchler, J. A. Genomic dosage effects on heterosis in triploid maize. *Proc. Natl Acad. Sci. USA* **110**, 2665–2669 (2013).
- Josephs, E. B., Lee, Y. W., Stinchcombe, J. R. & Wright, S. I. Association mapping reveals the role of purifying selection in the maintenance of genomic variation in gene expression. *Proc. Natl Acad. Sci. USA* **112**, 15390–15395 (2015).
- Gout, J.-F., Kahn, D., Duret, L. & Paramecium Post-Genomics Consortium. The relationship among gene expression, the evolution of gene dosage, and the rate of protein evolution. *PLoS Genet.* **6**, e1000944 (2010).
- Hufford, M. B. et al. Comparative population genomics of maize domestication and improvement. *Nat. Genet.* **44**, 808–811 (2012).
- Hung, H.-Y. et al. The relationship between parental genetic or phenotypic divergence and progeny variation in the maize nested association mapping population. *Heredity* **108**, 490–499 (2012).
- Rodgers-Melnick, E. et al. Recombination in diverse maize is stable, predictable, and associated with genetic load. *Proc. Natl Acad. Sci. USA* **112**, 3823–3828 (2015).

Supplementary Information is available in the online version of the paper.

Acknowledgements We thank J. Pardo, J. Wallace, R. Punna, K. Shirasawa and S. Miller for assistance with tissue collection; J. Budka and G. Inzinna for field and greenhouse assistance; R. Bukowski for running the maize HapMap genotyping pipeline; L. Johnson and Z. Miller for database curation; G. Gibson, M. Wolfe, J.-L. Jannink, M. Hufford and J. Ross-Ibarra for discussions; P. Schweitzer, J. Mosher, A. Tate, J. Mattison, M. Magallanes-Lundback, I. Holländer and D. Daujotyte for guidance on RNA extraction, library preparation automation and sequencing; and S. Miller for copy-editing. This work was supported by the US Department of Agriculture–Agricultural Research Service and the National Science Foundation grants IOS-0922493 and IOS-1238014 to E.S.B. The National Science Foundation Graduate Research Fellowship Program grant DGE-1650441 and the Section of Plant Breeding and Genetics at Cornell University provided support to K.A.G.K. The Taiwanese Ministry of Science and Technology Overseas Project for Post Graduate Research grant 104-2917-I-564-015 supported S.-Y.C.

Author Contributions K.A.G.K. and E.S.B. designed the experiments and wrote the manuscript. K.A.G.K. performed the analyses and made the RNA-seq libraries. K.A.G.K., S.-Y.C., and M.-H.S. extracted RNA. N.K.L. managed germplasm and plants with K.A.G.K., M.C.R., K.L.S. and A.L. produced and imputed HapMap genotypic data. P.J.B. implemented matrixEQL in Java/TASSEL. F.L. implemented SNP calling from RNA-seq data.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing interests. Readers are welcome to comment on the online version of the paper. Publisher's note: Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations. Correspondence and requests for materials should be addressed to E.S.B. (esb33@cornell.edu) and K.A.G.K. (kak268@cornell.edu).

Reviewer Information Nature thanks N. Springer and the other anonymous reviewer(s) for their contribution to the peer review of this work.

METHODS

No statistical methods were used to predetermine sample size. The investigators were not blinded to allocation during experiments and outcome assessment. Planting order was randomized for chamber and greenhouse experiments, as well as within maturity groups for field-grown plants. Flat order was also randomized.

Tissue collection and RNA extraction. Two centimetres of the base of leaf three from three uniform plants per genotype at third leaf stage were collected from 10:30 to 12:00. Tissue was collected 16–20 June 2015 and immediately frozen in liquid nitrogen. Seedlings were grown at one per cell in 96-cell flats in LM-111 mix (Lambert) at Guterma Greenhouse. Plants were watered daily and grown with 40–60% humidity under supplementary high-pressure 600-W sodium lighting from 8:30 to 20:30. Day temperature was 30 °C and night temperature was 24 °C. Flat positions were randomized daily.

Two centimetres of germinating seedling roots and whole germinating seedling shoots were collected from three plants per genotype from 11:00 to 13:00 on the day of germination and immediately frozen in liquid nitrogen. Seeds were germinated at eight per cell in 24-cell flats in medium grain vermiculite (Lambert) to facilitate collection of roots without soil contamination. A walk-in growth chamber was used with two 12-inch 24-W 6400K T5 fluorescent lights per shelf. Lights were on from 8:00 to 24:00. The lights were 36.5 cm from the shelves. Humidity ranged between 40 and 60%. Seeds were watered daily and the temperature was maintained between 22 and 24 °C. Flat positioning was randomized daily.

Seven developing kernels were collected from three plants at 350-growing-degree days after self-pollination and immediately frozen in liquid nitrogen. Based on pollination date, kernels were harvested between 11:00 and 13:00 from 25 August 2014 to 19 September 2014, from plants grown in field M1 (Lima silt loam soil) at the Cornell Musgrave Research Farm. Seven immature kernels were collected and combined from three plants per genotype. End plants from each plot were avoided whenever possible.

Using these same plots, mature leaves were collected during the day (11:00–13:00) and during the night (23:00–01:00) in two batches, on 8 August 2014 and 26 August 2014. Leaves were collected in two batches with the collection date assigned depending on what had already flowered by the dates above (see the column titled ‘Tissue’ in Supplementary Table 1 for collection date of adult leaves). Leaf sections from three plants per plot were collected into liquid nitrogen from a 1-cm section to one side of the midrib from the second leaf below the tassel. The proximal–distal middle of the leaf blade was determined by folding the leaf in half such that the leaf tip touched the ligule.

Non-kernel tissues from three plants per genotype \times tissue combination were homogenized with two 3-mm steel beads using a Genogrinder in 30-s increments to ensure samples remained frozen (Spex Sample Prep). Between each grinding increment, samples were placed in liquid nitrogen. RNA was extracted using TRIzol (Invitrogen) with Direct-zol columns (Zymo Research). Twenty-one frozen kernels (seven kernels from three individuals) per genotype were ground in an IKA seed mill (IKA). Hot borate and lithium chloride were used to extract RNA from kernels²⁹. For kernels, RNA integrity was assessed using gel electrophoresis and extractions were repeated for degraded samples. For all tissues, RNA was quantified using RNA Quantifluor (Promega) and diluted on a Beckman Biomek NXp to a concentration of 100 ng/ μ l for library preparation. 3' RNA-seq libraries were prepared robotically from 500 ng total RNA in 96-well plates on an NXp liquid handler (Beckman Coulter) using QuantSeq FWD kits (Lexogen) according to the manufacturer's instructions. Post-PCR cleanup was performed manually according to the QuantSeq protocol. Libraries were pooled to 96-plex, based on concentration as measured by DNA Quantifluor (Promega). Molar concentrations for each pool were calculated from Bioanalyzer (Agilent) fragment lengths and digital PCR quantifications. Pools were sequenced with 90 nucleotide single-end reads using Illumina TruSeq primers on an Illumina NextSeq 500 with v2 chemistry at the Cornell University Sequencing facility.

Trimomatic³⁰ version 0.32 was used to remove the first 12 bp and Illumina TruSeq adaptor remnants from each read. The first 12 bp were removed based on kit maker instructions, and the propensity for errors to occur in sequencing after random priming. The splice-aware STAR³¹ aligner v.2.4.2a was used to align reads against the maize genome annotation version AGPv.3.29, allowing a read to map in at most 10 locations (–outFilterMultimapNmax 10) with at most 4% mismatches (–outFilterMismatchNoverLmax 0.04), while filtering out all non-canonical intron motifs (–outFilterIntronMotifs RemoveNoncanonicalUnannotated). Default settings from HTSeq³² v.0.6.1 were used to obtain gene-level counts from the resulting BAM files. Counts were normalized by library sequencing depth using the –estimateSizeFactors method in DESeq2³³ in R. Expression counts were transformed using Box–Cox transformation after adding a small random value beneath the minimum detection threshold in order to enable transformation of zeroes.

Sample sizes. After positive sample identification by SNP-calling from RNA-seq data (as described below), sample sizes for each tissue were as follows: 2-cm root tips ($n = 291$, unique $n = 273$) and shoots ($n = 295$, unique $n = 278$) of germinating seedlings; 2-cm base ($n = 302$, unique $n = 263$) and tip ($n = 295$, unique $n = 265$) of leaf 3; 350-growing-degree-day kernels ($n = 254$, unique $n = 229$); and post-sexual-maturity (anthesis) leaves during the day ($n = 210$, unique $n = 204$) and night ($n = 276$, unique $n = 260$). Each sample was pooled from three individuals per genotype.

Genotyping. Genotypes from maize HapMap 3.2.1 were called as previously described²¹. Hapmap 3.2.1 was called on 1,268 inbred genotypes from across the world, with highly variable depth of coverage, and paralogous sites were retained, as they provide signal in genome-wide association studies. Because paralogous sites were retained, we used k -nearest neighbour imputation³⁴ (KNNi) as implemented in TASSEL³⁵ to impute; this was robust to high error rates in genotype calling, but KNNi over-imputes missing data to the major allele. Despite this bias, KNNi imputation for HapMap 3.2.1 had an overall accuracy of 0.988 and a minor allele imputation accuracy of 0.94 for imputed genotypes. Imputed HapMap 3.2.1 genotypes were projected onto the genotyping by sequencing (GBS)-genotyped inbred diversity panel (including tropical lines²²) using FILLIN³⁶. Projection was anchored by 465,085 consensus sites between HapMap3 and GBS, in which the physical positions match and the major/minor alleles are shared; the projection accuracy was $r = 0.99$ between masked and subsequently imputed genotypes (0.96 for minor alleles).

To rapidly confirm the identity of the RNA-seq samples, SNPs were called from RNA-seq reads using FastCall³⁷. Using SNPs called from the first 20 Mb on chromosome 10 (AGPv.3), all samples that did not match the reference genotype were discarded, leaving 1,960 samples. These SNPs called from RNA-seq were not used for eQTL association studies or other downstream analyses.

Gene set filtering. After library size normalization with DESeq2 (described above), genes were ranked by expression in each tissue and the top 5,000 were placed in the ‘top 5,000’ set; the next 5,000 were placed in the ‘5,001–10,000’ set for each tissue. The blocks were further broken down into groups of 1,000 for the plots included in the Extended Data.

eQTL mapping and 2D MAF comparisons between tropical and RNA sets. Before mapping eQTL, 25 hidden factors were calculated using PEER³⁸ for each tissue individually and these were used as covariates together with 5 multidimensional scaling (MDS) coordinates calculated from HapMap 3.2.1²¹. Using the PEER factors and MDS values, eQTL were mapped using a parallelized version of MatrixEQTL³⁹ implemented in TASSEL using Java. P values and R^2 from eQTL hits with significance below $P = 0.00001$ were recorded after performing association tests individually for each tissue with SNPs that had a MAF ≥ 0.05 in the RNA-set samples for a specific tissue.

Each SNP present in the temperate-biased RNA set and previously published tropical lines²² was then plotted in a 2D minor-allele frequency plot, with its position on the plot determined by its frequency in the tropical and predominately temperate RNA set. Recorded R^2 from *cis* eQTL (within ± 1 Mb) were then used to colour each dot, revealing the effects of ancestral allele frequencies on modern eQTL in Fig. 2. Within each tissue, the most strongly associated SNP–gene pair was kept for each SNP for plotting in the 2D MAF plots.

The distribution of variance explained by SNPs that were formerly rare (tropical MAF under 0.1), but are now common (RNA-set MAF over 0.4) was compared to SNPs that were common in the tropical populations and remain common in the RNA set, which revealed that formerly rare SNPs explain more variance.

Calculation of seed-weight fitness. Multi-location multi-year seed weight best linear unbiased predictions²⁷ were used as a proxy for fitness. Seed weights from sweet corn and popcorn lines were excluded from Extended Data Fig. 8 and Extended Data Table 1, given that dry seed weight is not a rational proxy for fitness or target of selection in these subpopulations. To be conservative, tropical lines were also excluded from Fig. 3 and Extended Data Fig. 7 because they are likely to be substantially dysregulated relative to the temperate material and have lower seed weights, thus inflating prediction accuracy of fitness from dysregulation. However, significant prediction accuracy is still achieved when tropical lines are included (data not shown).

Prediction of fitness from dysregulation of expression. To quantify the deviation of expression, we subtracted the mean expression level for a single gene from the expression of that gene in each line. This was done for the 5,000 highest expressed genes in each tissue. These 5,000 deviations were then used as X values in a ridge regression ($\alpha = 0$) implemented using the glmnet package⁴⁰ in R to predict the seed weight best linear unbiased predictions, discussed above. Ten repetitions of nested tenfold (tenfold inner, tenfold outer) cross-validation were carried out using glmnet. In Extended Data Fig. 8, the cumulative absolute deviation in expression for the top 5,000 most-expressed genes (see formula below) was also correlated against the seed weights.

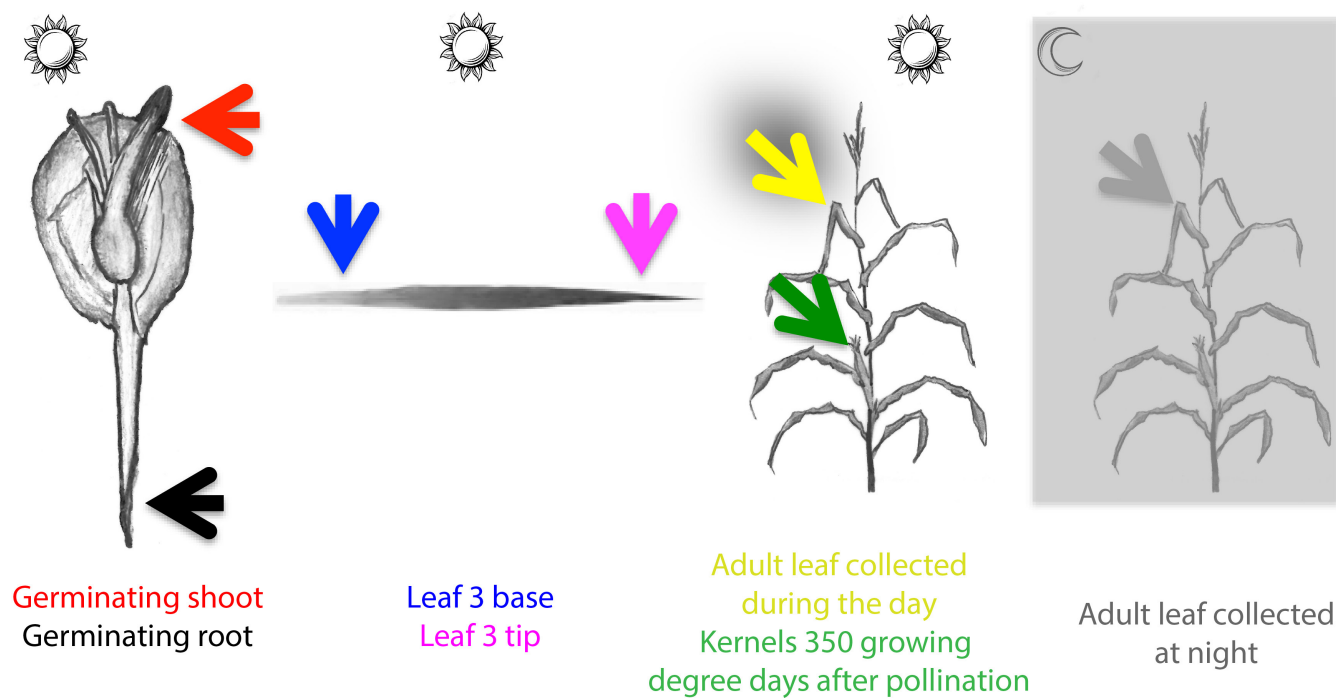
$$D_i = \log_{10} \frac{\sum_{j=1}^n (e_{ij} - \mu_j)^2}{n}$$

in which D_i is the deviation of expression of individual i , e_{ij} is the expression of gene j in individual i , n is the number of genes and μ_j is the mean expression of gene j across all samples profiled in the tissue.

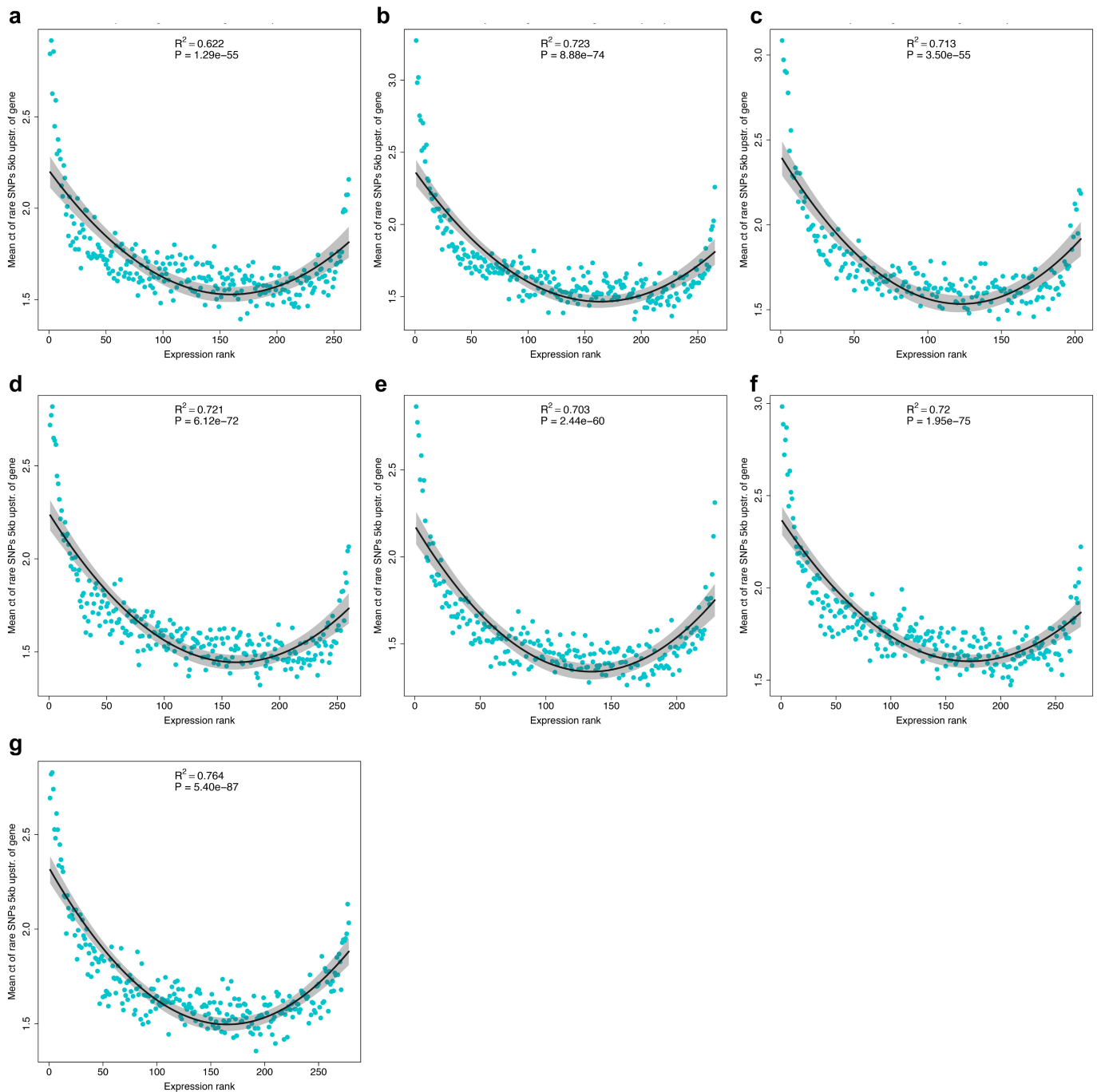
Code availability. Robotic code for Biomek NXp RNA-seq library production can be found at <http://www.maizegenetics.net/robotic-code>; parallelized implementation of matrix eQTL in TASSEL can be found at <http://www.maizegenetics.net/tassel>; and FastCall for calling SNPs from whole genome sequencing data can be found at <https://github.com/Fei-Lu/FastCall>. Further scripts for making plots are available from the corresponding authors upon reasonable request.

Data availability. Sequence data that support the findings of this study have been deposited in the Sequence Read Archive under accession number SRP115041, and in BioProject under accession number PRJNA383416. Processed expression counts are available at the Cyverse Discovery Environment (de.cyverse.org/de/) under the directory: /iplant/home/shared/panzea/dataFromPubs/. All other data are available from the corresponding author(s) upon reasonable request.

29. Wan, C. Y. & Wilkins, T. A. A modified hot borate method significantly enhances the yield of high-quality RNA from cotton (*Gossypium hirsutum* L.). *Anal. Biochem.* **223**, 7–12 (1994).
30. Bolger, A. M., Lohse, M. & Usadel, B. Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* **30**, 2114–2120 (2014).
31. Dobin, A. *et al.* STAR: ultrafast universal RNA-seq aligner. *Bioinformatics* **29**, 15–21 (2013).
32. Anders, S., Pyl, P. T. & Huber, W. HTSeq—a Python framework to work with high-throughput sequencing data. *Bioinformatics* **31**, 166–169 (2015).
33. Anders, S. & Huber, W. Differential expression analysis for sequence count data. *Genome Biol.* **11**, R106 (2010).
34. Money, D. *et al.* LinkImpute: fast and accurate genotype imputation for nonmodel organisms. *G3* **5**, 2383–2390 (2015).
35. Bradbury, P. J. *et al.* TASSEL: software for association mapping of complex traits in diverse samples. *Bioinformatics* **23**, 2633–2635 (2007).
36. Swarts, K. *et al.* Novel methods to optimize genotypic imputation for low-coverage, next-generation sequence data in crop plants. *Plant Genome* **7**, <http://doi.org/10.3835/plantgenome2014.05.0023> (2014).
37. Ramu, P. *et al.* Cassava haplotype map highlights fixation of deleterious mutations during clonal propagation. *Nat. Genet.* **49**, 959–963 (2017).
38. Stegle, O., Parts, L., Durbin, R. & Winn, J. A Bayesian framework to account for complex non-genetic factors in gene expression levels greatly increases power in eQTL studies. *PLOS Comput. Biol.* **6**, e1000770 (2010).
39. Shabalin, A. A. Matrix eQTL: ultra fast eQTL analysis via large matrix operations. *Bioinformatics* **28**, 1353–1358 (2012).
40. Friedman, J., Hastie, T. & Tibshirani, R. Regularization paths for generalized linear models via coordinate descent. *J. Stat. Softw.* **33**, 1–22 (2010).
41. Kisselbach, T. A. *The Structure and Reproduction of Corn* (Cold Spring Harbor Laboratory, 1999).

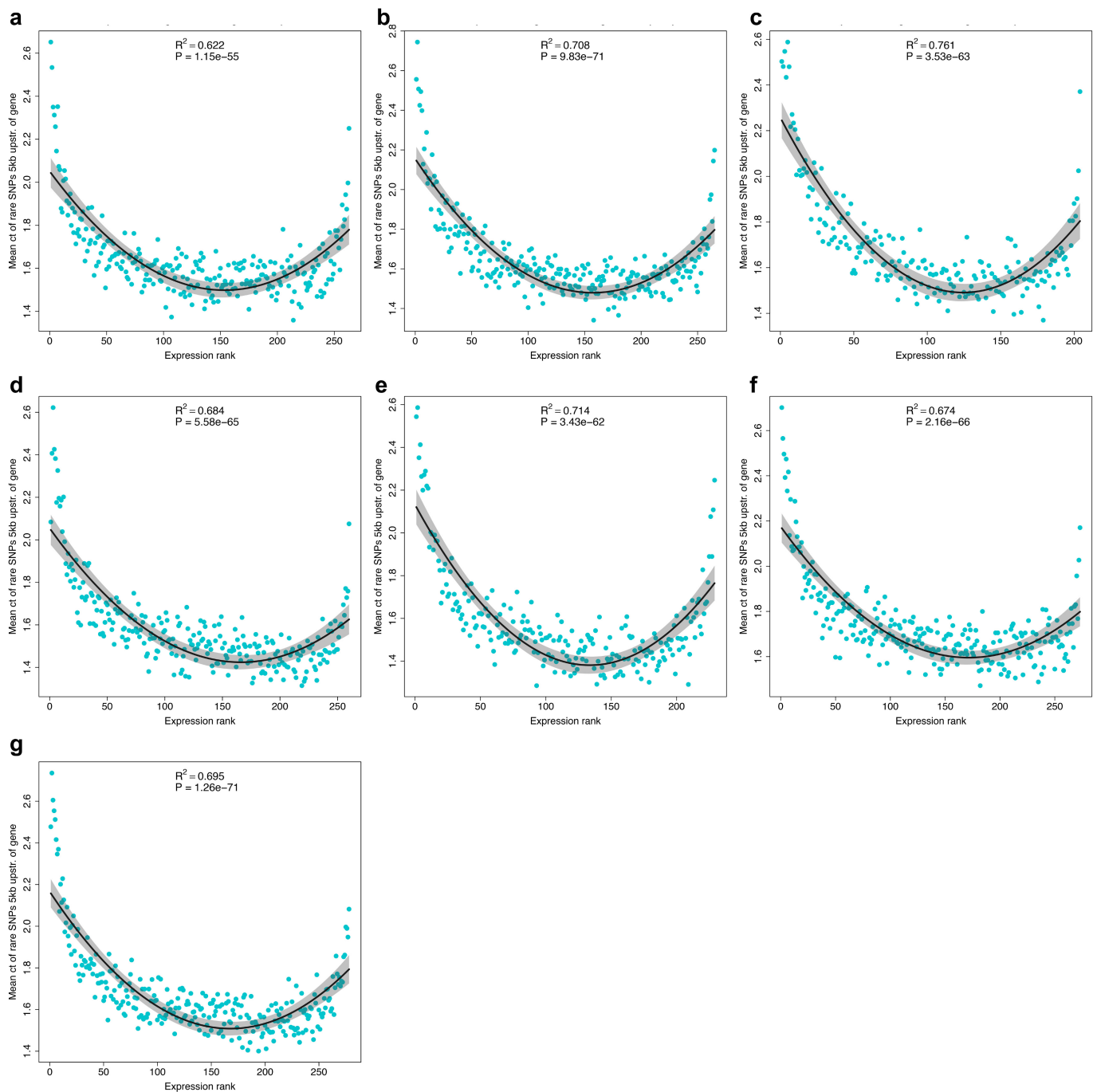


Extended Data Figure 1 | Tissues that were expression profiled by 3' RNA-seq. See additional details regarding tissue collection in Methods. Illustrations inspired by ref. 41.



Extended Data Figure 2 | Higher numbers of rare alleles are upstream of genes in extreme-expressing individuals, for the most highly expressed genes. Quadratic regression of the expression rank of each line, for each of the top 5,000 most-expressed genes versus the average local (5-kb upstream) rare-allele count. **a**, Base of leaf three ($n = 263$ unique inbred samples). **b**, Tip of leaf three ($n = 265$ unique inbred samples).

c, Adult leaves collected during the day ($n = 204$ unique inbred samples). **d**, Adult leaves collected at night ($n = 260$ unique inbred samples). **e**, Kernels at 350-growing-degree days ($n = 229$ unique inbred samples). **f**, Roots of germinating seedling ($n = 273$ unique inbred samples). **g**, Shoots of germinating seedling ($n = 278$ unique inbred samples).



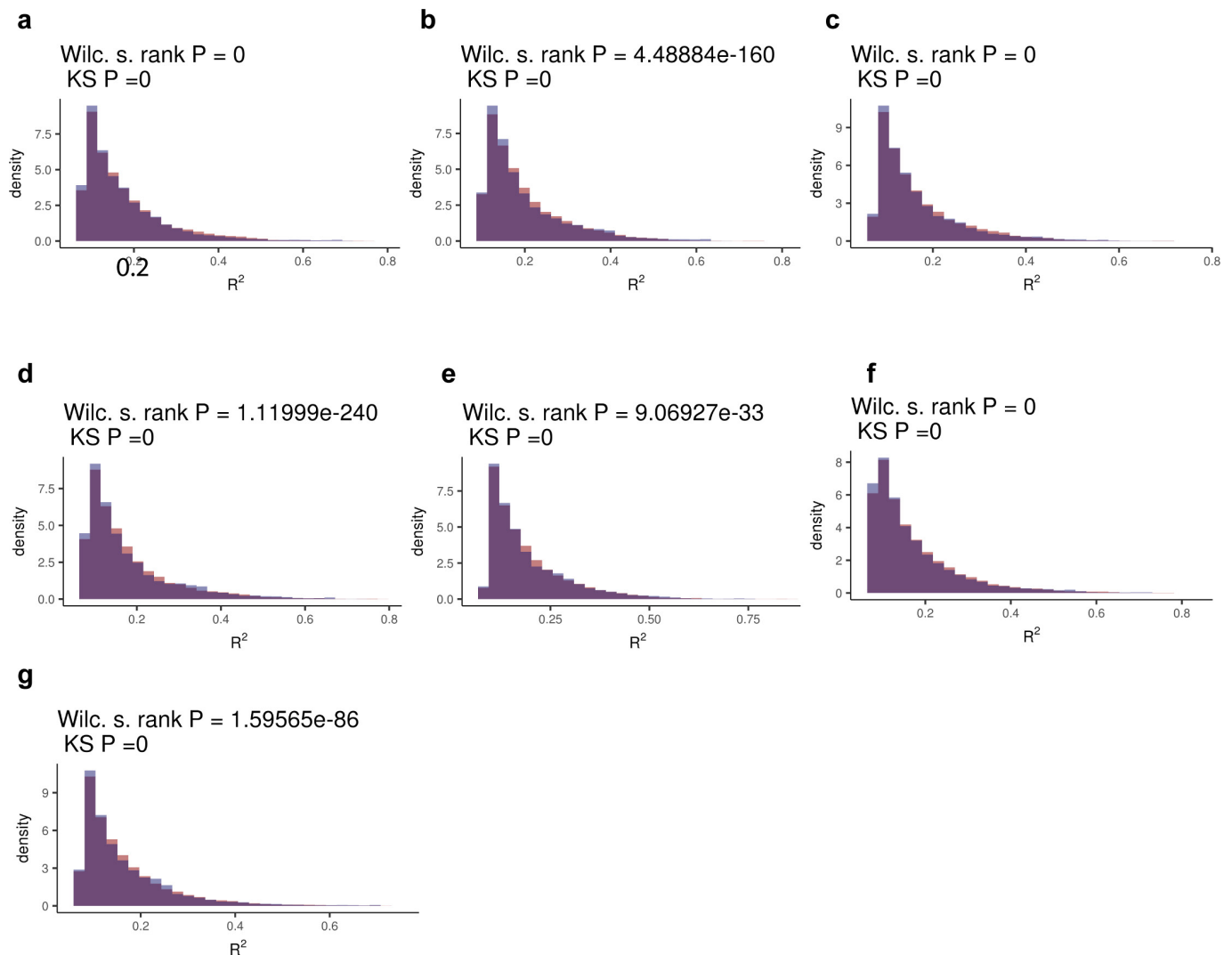
Extended Data Figure 3 | Higher numbers of rare alleles are upstream of genes in extreme-expressing individuals, for the medium-expressed genes. Quadratic regression of the expression rank of each line, for each of the top 5,001–10,000 most-expressed genes versus the average local (5-kb upstream) rare-allele count. **a**, Base of leaf three ($n = 263$ unique inbred samples). **b**, Tip of leaf three ($n = 265$ unique inbred samples). **c**, Adult

leaves collected during the day ($n = 204$ unique inbred samples). **d**, Adult leaves collected at night ($n = 260$ unique inbred samples). **e**, Kernels at 350-growing-degree days ($n = 229$ unique inbred samples). **f**, Roots of germinating seedling ($n = 273$ unique inbred samples). **g**, Shoots of germinating seedling ($n = 278$ unique inbred samples).



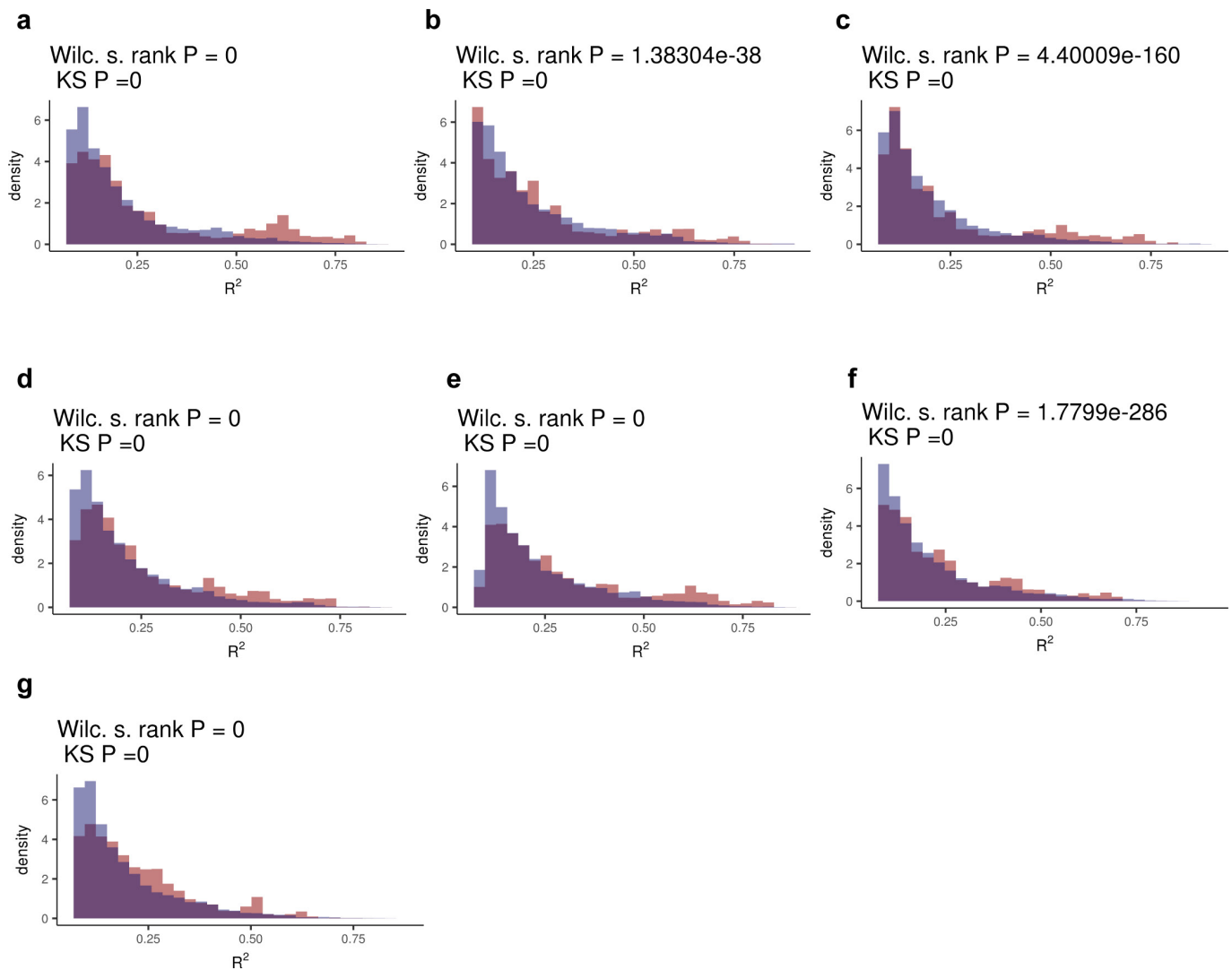
Extended Data Figure 4 | Comparison of the number of rare *cis* alleles near genes with differing expression levels. The 10,000 most-expressed genes in each tissue are divided into groups of 1,000 on the basis of expression level. Plots in each panel show genes ranked 1–1,000, 1,001–2,000, ..., 9,001–10,000 from left to right. Each of the individuals represented in each tissue is ranked for expression for each of the 1,000 genes in each group. Individuals in the bottom five expression ranks (fuchsia) versus the middle two quartiles (yellow) versus the top five

expression ranks (blue) (mean \pm s.e.m.). Y axes refer to mean upstream (within 5 kb) rare-allele count. **a**, Roots of germinating seedling ($n = 273$ unique inbred samples). **b**, Shoots of germinating seedling ($n = 278$ unique inbred samples). **c**, Kernels at 350-growing-degree days ($n = 229$ unique inbred samples). **d**, Base of leaf three ($n = 263$ unique inbred samples). **e**, Tip of leaf three ($n = 265$ unique inbred samples). **f**, Adult leaves collected during the day ($n = 204$ unique inbred samples). **g**, Adult leaves collected at night ($n = 260$ unique inbred samples).



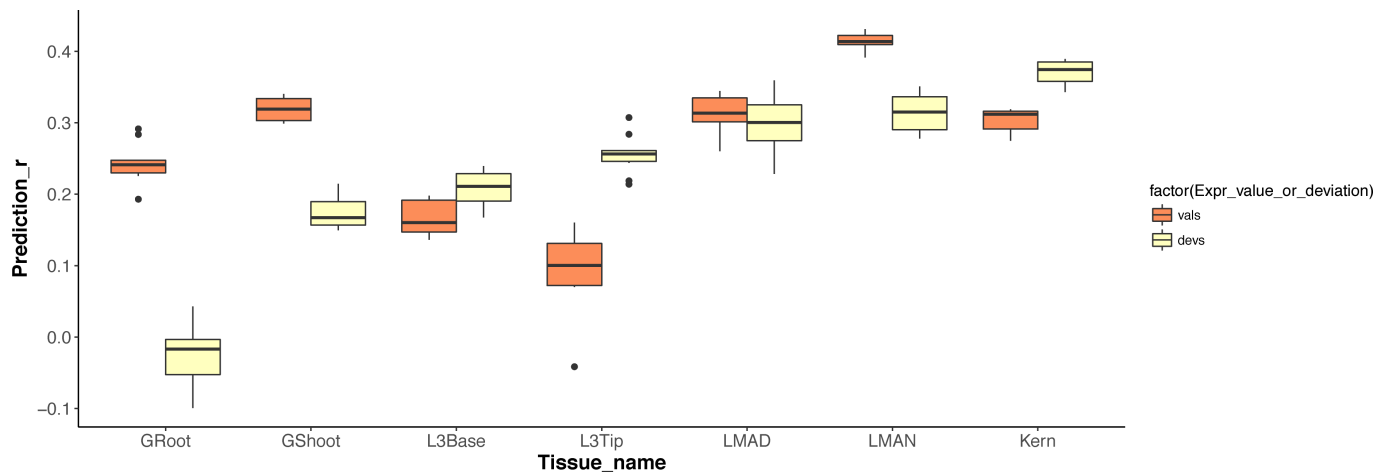
Extended Data Figure 5 | eQTL R^2 distribution comparisons between SNPs in 0.0–0.1 (tropical MAF) and 0.1–0.2 (RNA-set MAF) versus 0.1–0.2 (RNA-set and tropical MAF). **a**, Adult leaves collected at night ($n = 260$ unique inbred samples). **b**, Adult leaves collected during the day ($n = 204$ unique inbred samples). **c**, Tip of leaf three ($n = 265$ unique inbred samples). **d**, Base of leaf three ($n = 263$ unique inbred samples).

e, Kernels at 350-growing-degree days ($n = 229$ unique inbred samples). **f**, Shoots of germinating seedling ($n = 278$ unique inbred samples). **g**, Roots of germinating seedling ($n = 273$ unique inbred samples). All pairs of distributions within each tissue are significantly different. $P < 2.2 \times 10^{-16}$ two-sided Wilcoxon signed-rank test and Kolmogorov–Smirnov test.



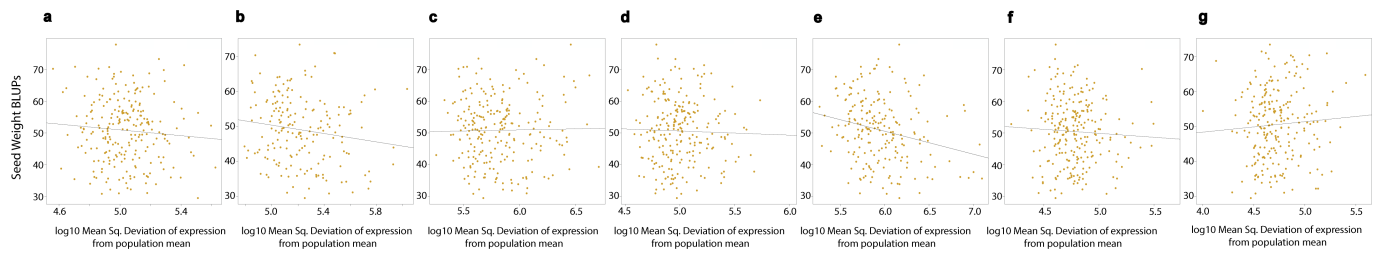
Extended Data Figure 6 | eQTL R^2 distribution comparisons between SNPs in 0.0–0.1 (tropical MAF) and 0.4–0.5 (RNA-set MAF) versus 0.4–0.5 (RNA-set and tropical MAF). **a**, Adult leaves collected at night ($n = 260$ unique inbred samples). **b**, Adult leaves collected during the day ($n = 204$ unique inbred samples). **c**, Tip of leaf three ($n = 265$ unique inbred samples). **d**, Base of leaf three ($n = 263$ unique inbred samples).

e, Kernels at 350-growing-degree days ($n = 229$ unique inbred samples). **f**, Shoots of germinating seedling ($n = 278$ unique inbred samples). **g**, Roots of germinating seedling ($n = 273$ unique inbred samples). All pairs of distributions within each tissue are significantly different. $P < 2.2 \times 10^{-16}$ two-sided Wilcoxon signed-rank test and Kolmogorov–Smirnov test.



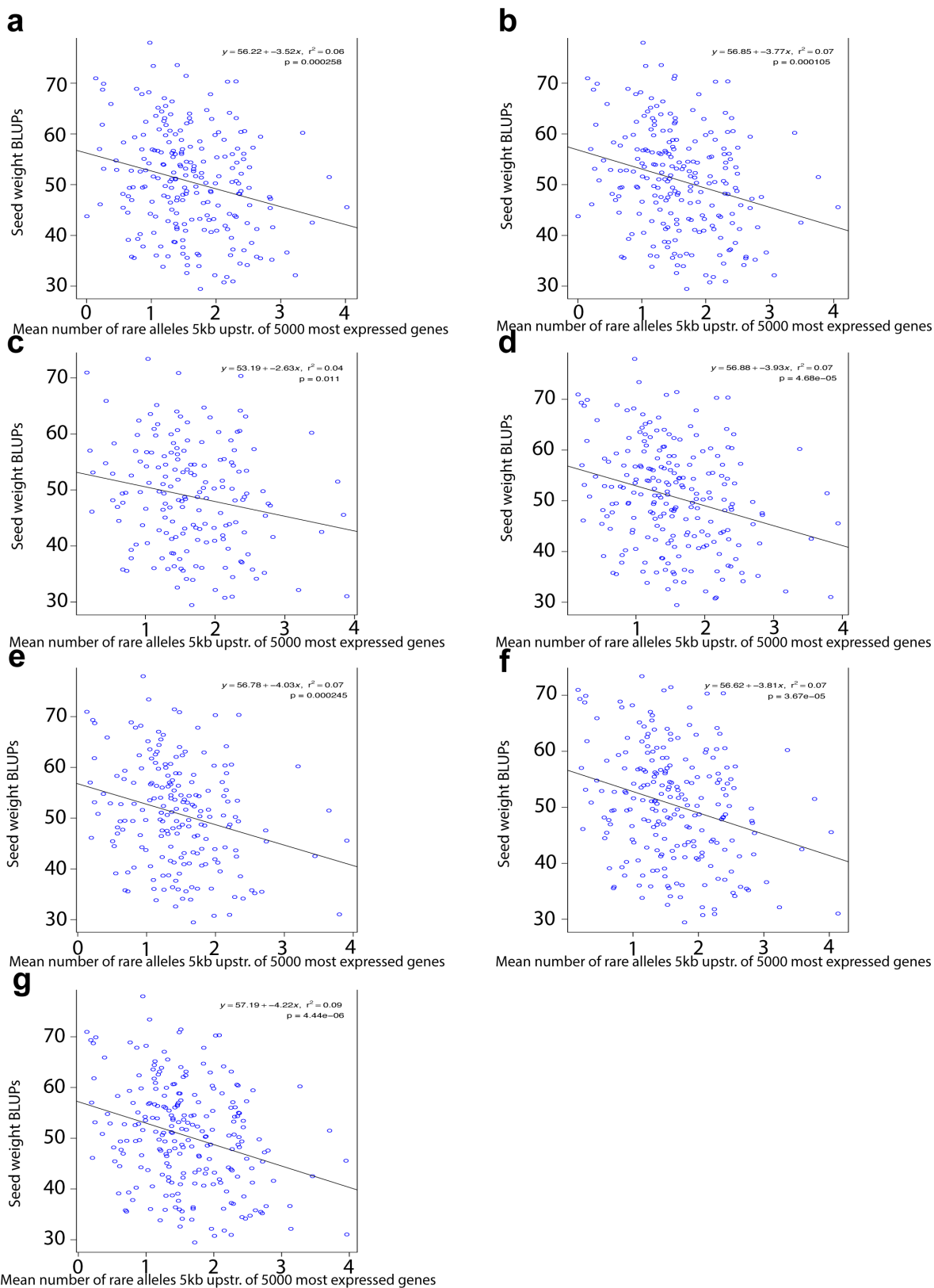
Extended Data Figure 7 | Expression value and dysregulation of 5,000 most-expressed genes are both predictive of fitness. Orange boxes represent correlations between predicted and true seed weight when using expression values. Yellow boxes represent correlations between predicted and true seed weight when using absolute deviation in expression from the population mean. Range of correlations between predicted and true seed weight is displayed from ten repetitions of nested tenfold cross validation (ten inner and ten outer) using ridge regression. In the box plots, the middle horizontal lines represent the median, hinges represent the

25th and 75th percentiles (the interquartile range), the upper and lower whiskers extend to maximum and minimum points no more than $1.5 \times$ interquartile range beyond the hinges, and individual dots are outliers beyond the whiskers. Sample sizes: 2-cm root tips of germinating seedlings (unique $n = 181$) and whole shoots of germinating seedlings (unique $n = 183$); the 2-cm base (unique $n = 181$) and tip (unique $n = 182$) of leaf 3; leaves collected in the field during the day (unique $n = 135$) and night (unique $n = 187$); and 350-growing-degree-day kernels (unique $n = 171$), post sexual maturity (anthesis).



Extended Data Figure 8 | Cumulative expression dysregulation of the 5,000 most-expressed genes in each tissue versus seed weight. a, Adult leaves collected at night ($n = 221$ unique inbred samples). **b,** Adult leaves collected during the day ($n = 171$ unique inbred samples). **c,** Tip of leaf three ($n = 226$ unique inbred samples). **d,** Base of leaf three ($n = 224$

unique inbred samples). **e,** Kernels at 350-growing-degree days ($n = 195$ unique inbred samples). **f,** Shoots of germinating seedling ($n = 235$ unique inbred samples). **g,** Roots of germinating seedling ($n = 226$ unique inbred samples). Regression statistics in Extended Data Table 1. Sweet corn and popcorn lines were excluded from these regressions.



Extended Data Figure 9 | Mean upstream rare-allele count from the 5,000 most highly expressed genes versus seed weight. **a**, Adult leaves collected at night ($n = 221$ unique inbred samples). **b**, Adult leaves collected during the day ($n = 171$ unique inbred samples). **c**, Tip of leaf three ($n = 226$ unique inbred samples). **d**, Base of leaf three ($n = 224$

unique inbred samples). **e**, Kernels at 350-growing-degree days ($n = 195$ unique inbred samples). **f**, Shoots of germinating seedling ($n = 235$ unique inbred samples). **g**, Roots of germinating seedling ($n = 226$ unique inbred samples).

Extended Data Table 1 | Regression statistics for cumulative expression dysregulation in each tissue against seed-weight fitness

| Tissue | P-value | r | n |
|----------------------|---------|--------|-----|
| Kernel | 0.00046 | -0.242 | 195 |
| Adult Leaves - day | 0.03886 | -0.154 | 171 |
| Germinating Shoot | 0.42226 | -0.052 | 235 |
| Adult Leaves - night | 0.20301 | -0.084 | 221 |
| Germinating root | 0.29569 | 0.069 | 226 |
| Leaf 3 base | 0.66049 | -0.029 | 224 |
| Leaf 3 tip | 0.77367 | 0.019 | 226 |

Sample size *n* refers to genetically unique inbred samples after excluding sweet corn and popcorn lines.

A single-cell RNA-seq survey of the developmental landscape of the human prefrontal cortex

Suijuan Zhong^{1,2*}, Shu Zhang^{3*}, Xiaoying Fan^{3*}, Qian Wu^{1,2*}, Liying Yan^{3*}, Ji Dong³, Haofeng Zhang⁴, Long Li^{1,2}, Le Sun¹, Na Pan¹, Xiaohui Xu⁴, Fuchou Tang^{3,5,6}, Jun Zhang⁴, Jie Qiao^{3,5,6} & Xiaoqun Wang^{1,2,7}

The mammalian prefrontal cortex comprises a set of highly specialized brain areas containing billions of cells and serves as the centre of the highest-order cognitive functions, such as memory, cognitive ability, decision-making and social behaviour^{1,2}. Although neural circuits are formed in the late stages of human embryonic development and even after birth, diverse classes of functional cells are generated and migrate to the appropriate locations earlier in development. Dysfunction of the prefrontal cortex contributes to cognitive deficits and the majority of neurodevelopmental disorders; there is therefore a need for detailed knowledge of the development of the prefrontal cortex. However, it is still difficult to identify cell types in the developing human prefrontal cortex and to distinguish their developmental features. Here we analyse more than 2,300 single cells in the developing human prefrontal cortex from gestational weeks 8 to 26 using RNA sequencing. We identify 35 subtypes of cells in six main classes and trace the developmental trajectories of these cells. Detailed analysis of neural progenitor cells highlights new marker genes and unique developmental features of intermediate progenitor cells. We also map the timeline of neurogenesis of excitatory neurons in the prefrontal cortex and detect the presence of interneuron progenitors in early developing prefrontal cortex. Moreover, we reveal the intrinsic development-dependent signals that regulate neuron generation and circuit formation using single-cell transcriptomic data analysis. Our screening and characterization approach provides a blueprint for understanding the development of the human prefrontal cortex in the early and mid-gestational stages in order to systematically dissect the cellular basis and molecular regulation of prefrontal cortex function in humans.

The prefrontal cortex (PFC) covers the front of the frontal lobe of the brain in mammals and has important roles in memory, emotion, cognitive behaviour, decision-making and social behaviour^{1–4}. To analyse the molecular features of cells in the PFC during human brain development, we obtained 2,309 single cells from human embryonic PFCs at gestational weeks (GW)8 to 26 (specifically, GW8, GW9, GW10, GW12, GW13, GW16, GW19, GW23 and GW26; three and two biological replicates at GW10 and GW23, respectively; Extended Data Fig. 1a, b). To classify the major cell types in the developing PFC, we performed *t*-distributed stochastic neighbour embedding (*t*-SNE) analysis using Seurat⁵ and identified six major clusters: neural progenitor cells (NPCs), excitatory neurons, interneurons, astrocytes, oligodendrocyte progenitor cells (OPCs) and microglia (Fig. 1a, Extended Data Figs 1c, d, 2a, b). Biological replicate samples were evenly distributed on the *t*-SNE plot (Extended Data Fig. 1c). To further analyse the subclusters within each cell type, we used random forest⁶ analysis to segregate cells into 35 distinct subtypes (Fig. 1b, Extended

Data Fig. 2c). By analysing differentially expressed genes among the clusters (Fig. 1b, Supplementary Table 1), we identified *SFRP1* and *RBFOX1* as markers of NPCs and excitatory neurons and verified this by immunofluorescence (Extended Data Fig. 2d). We found that NPCs, excitatory neurons and interneurons were sub-clustered in a development-dependent manner (Extended Data Fig. 3). We then reconstructed the developmental time course and lineage relationships using Monocle analysis^{7–9}. Microglia and interneurons were excluded because microglia are mesoderm-derived cells⁷ and interneurons are generated in the ganglionic eminence and migrate tangentially to the PFC^{10–12}. The remaining cells were distributed along pseudo-temporally ordered paths from NPCs to neurons (neuronal lineage) or to OPCs and astrocytes (glial lineage) (Fig. 1c, Extended Data Fig. 4a). Neurons developed from NPCs in early gestational weeks whereas OPCs and astrocytes differentiated from NPCs in later weeks (Fig. 1d, Extended Data Fig. 4b).

Using pseudo-time analysis, we found decreased SOX2 expression in the neuronal lineage but sustained high expression in the glial lineage (Fig. 1e, Extended Data Fig. 4). Next, we performed immunofluorescence staining of PAX6, SOX2 and NEUROD2 in embryos from GW8 to GW23. In the PFC from GW8 embryos, the majority of PAX6⁺SOX2⁺ cells were tightly localized in the ventricular zone. At GW12, the subventricular zone (SVZ) included a mixture of NPCs and a few neurons. As the PFC developed, the ventricular zone narrowed, whereas the NPC-enriched outer SVZ (oSVZ) expanded at GW16. At GW23, many of the SOX2⁺ cells in the oSVZ did not express PAX6 (Extended Data Fig. 5a, b), which suggests that the progenitor cells had developed into other lineages, such as oligodendrocytes and astrocytes, in the PFC, consistent with single-cell analysis showing that the majority of OPCs and astrocytes appeared in GW23 and GW26.

Progenitors of microglia arise from peripheral mesodermal tissue, but not from the neuroectoderm¹³. We observed that microglia appeared early in the PFC and were present throughout PFC development (Fig. 1a). Microglia were present outside the ventricular zone at GW8, when they had amoeboid or star-like morphology (Fig. 1f, Extended Data Fig. 5c). Microglia migrated to the SVZ and intermediate zone at GW12 and GW16, and then penetrated to the ventricular zone and cortical plate at GW19 and GW23 (Fig. 1f, g, Extended Data Fig. 5d). Microglia in the intermediate zone exhibited multi-directionally orientated processes, and those in the ventricular zone and SVZ had a simpler morphology at GW19 (Fig. 1f). The early appearance and stable population of microglia in the developing PFC, especially in the intermediate zone, support a model in which microglia are involved in regulating neuronal apoptosis, neurogenesis and synaptic pruning in development^{14–16}.

To reveal the diversity of NPCs in human PFC, we identified nine subclusters by gene ontology analysis of the differentially expressed

¹State Key Laboratory of Brain and Cognitive Science, CAS Center for Excellence in Brain Science and Intelligence Technology (Shanghai), Institute of Biophysics, Chinese Academy of Sciences, Beijing, 100101, China. ²University of Chinese Academy of Sciences, Beijing, 100049, China. ³Beijing Advanced Innovation Center for Genomics, College of Life Sciences, Department of Obstetrics and Gynecology, Third Hospital, Peking University, Beijing, 100871, China. ⁴Obstetrics and Gynecology, Medical Center of Severe Cardiovascular of Beijing Anzhen Hospital, Capital Medical University, Beijing, 100029, China. ⁵Biomedical Institute for Pioneering Investigation via Convergence and Center for Reproductive Medicine, Ministry of Education Key Laboratory of Cell Proliferation and Differentiation, Beijing, 100871, China. ⁶Peking-Tsinghua Center for Life Sciences, Peking University, Beijing, 100871, China. ⁷Beijing Institute for Brain Disorders, Beijing, 100069, China.

*These authors contributed equally to this work.

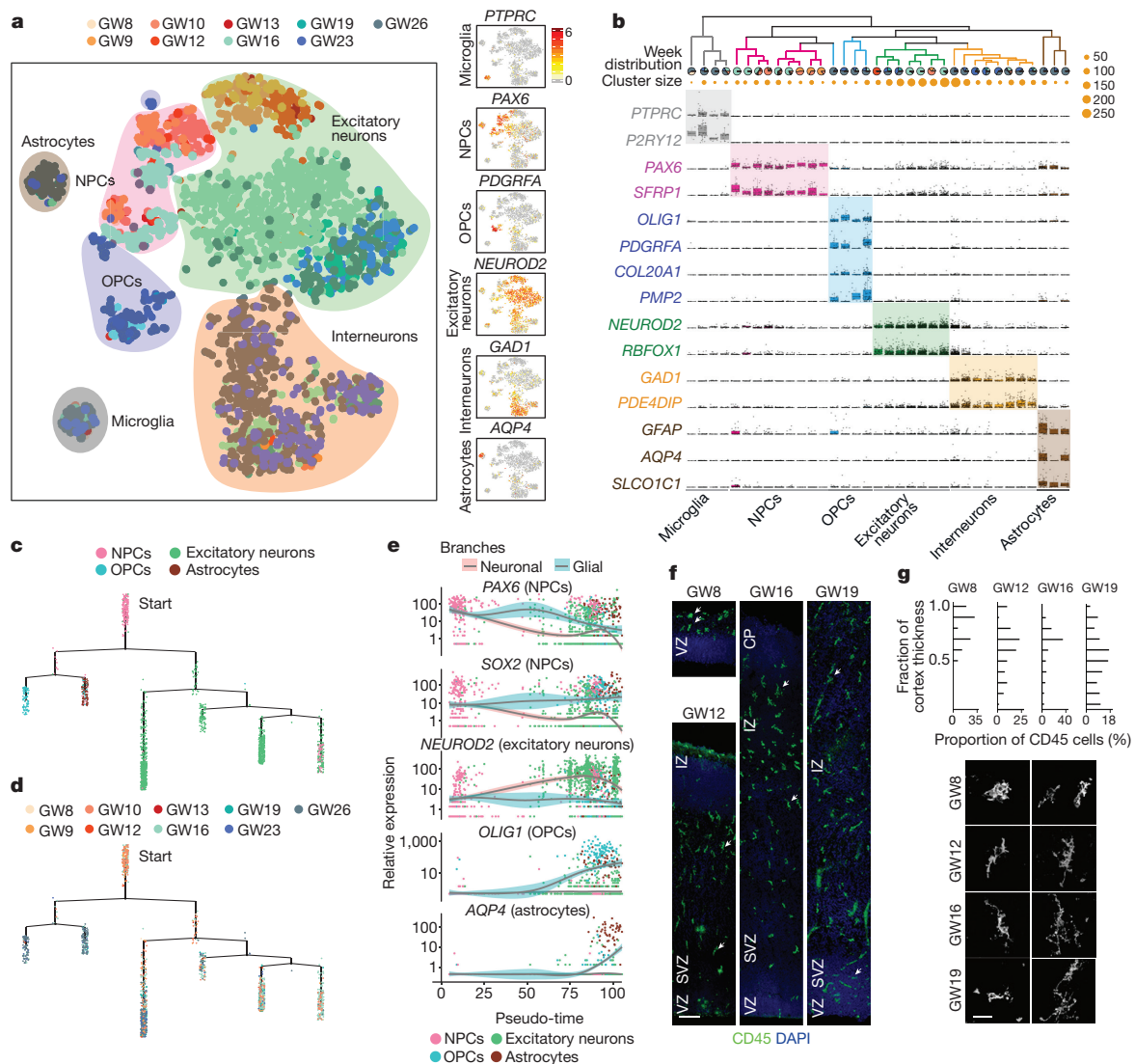


Figure 1 | Molecular diversity of cells from the developing human PFC.

a, Visualization of major classes of cells using *t*-SNE. Dots, individual cells; colour, gestational weeks (GW); colour contours, cell types. Right, expression of known markers; grey, no expression; yellow–red, relative expression. **b**, Hierarchical clustering analysis of 35 subclasses. Microglia: $n = 5, 38, 10$ and 15 cells, left to right; NPCs: $n = 28, 23, 44, 60, 54, 21, 15, 38$ and 7 cells, left to right; OPCs: $n = 22, 28, 12$ and 55 cells, left to right; excitatory neurons: $n = 53, 103, 149, 163, 189, 142$ and 258 cells, left to right; interneurons: $n = 261, 120, 37, 64, 24, 63, 80$ and 52 cells, left to right; astrocytes: $n = 47, 12$ and 17 cells, left to right. Pie charts, distribution over gestational weeks; orange dots, subclass size. **c**, **d**, Single-cell trajectories by Monocle analysis showing development of the PFC.

genes (Fig. 2a, Extended Data Fig. 6a). *HOPX*, *FAM107A* and *TNC*, which have been identified as markers of outer radial glia (oRG)^{17,18}, were expressed in clusters 7, 8 and 9, while *EOMES* was expressed in clusters 3, 4 and 5, which suggests that the NPCs in these clusters were intermediate progenitor cells (IPCs) (Extended Data Fig. 6b). Using Monocle analysis, we found that the developmental trajectory of NPCs followed three major paths: early and late paths from ventricular radial glia (vRG) to IPC and a late path to oRG cells (Fig. 2b). To identify genes regulating radial glial cells in symmetric, neurogenic or gliogenic divisions¹⁹, we analysed the differentially expressed genes in *EOMES*[−] NPCs from GW8 to GW10 and from GW16 to GW19 (Extended Data Fig. 6c). We found that *HMG2* was expressed in *SOX2*⁺ and *Ki-67*⁺ cells in the ventricular zone and SVZ of GW10 PFC but not in NPCs from GW16 PFC (Extended Data Fig. 6d). In addition, several differentially expressed genes in *EOMES*[−] NPCs

The distance from a cell to the root corresponds to pseudo-time. Branched trajectories are plotted as a two-dimensional tree layout. **e**, Expression of known markers with pseudo-time. The lines with pink and blue shadows represent the neuronal and glial branches, respectively. The shadow represents the confidence interval around the fitted curve. $n = 1,540$ cells. **f**, The distribution and morphology of microglia in the PFC. Scale bars, $100\ \mu\text{m}$ (left) and $20\ \mu\text{m}$ (right). **g**, Quantification of the distribution of microglia in **f**. CD45⁺ cells were distributed across ten evenly spaced bins from the ventricular zone to the cortical plate. The experiment was repeated three times independently with similar results. See Extended Data Figs 1–5.

from GW16 to GW19 were oRG markers, and some were known neuronal markers (Extended Data Fig. 6c). Therefore, we co-stained GW16 PFC for *HOPX*, *SOX2* and *NEUROD1*, and observed that some *HOPX*⁺*SOX2*⁺ cells also expressed *NEUROD1* in oSVZ (Extended Data Fig. 6d), indicating that these progenitors might undergo neurogenic cell division and the resulting neurons may retain markers from the progenitor cells. Consistent with this hypothesis, the gene ontology results indicated that the oRG cells in cluster 8 (GW9 to GW16) were actively proliferating, whereas the oRG cells in cluster 7 (GW16 to GW26) were undergoing neuronal differentiation (Extended Data Fig. 6a).

IPCs act as transient amplifying progenitor cells, enabling the rapid generation of neurons during cortical development^{20–23}. To illustrate IPC development, we assessed the relative expression of *EOMES* in pseudo-time (Fig. 2b, c) and the proportion of *EOMES*⁺ cells at different gestational times, and found that they peaked during GW10

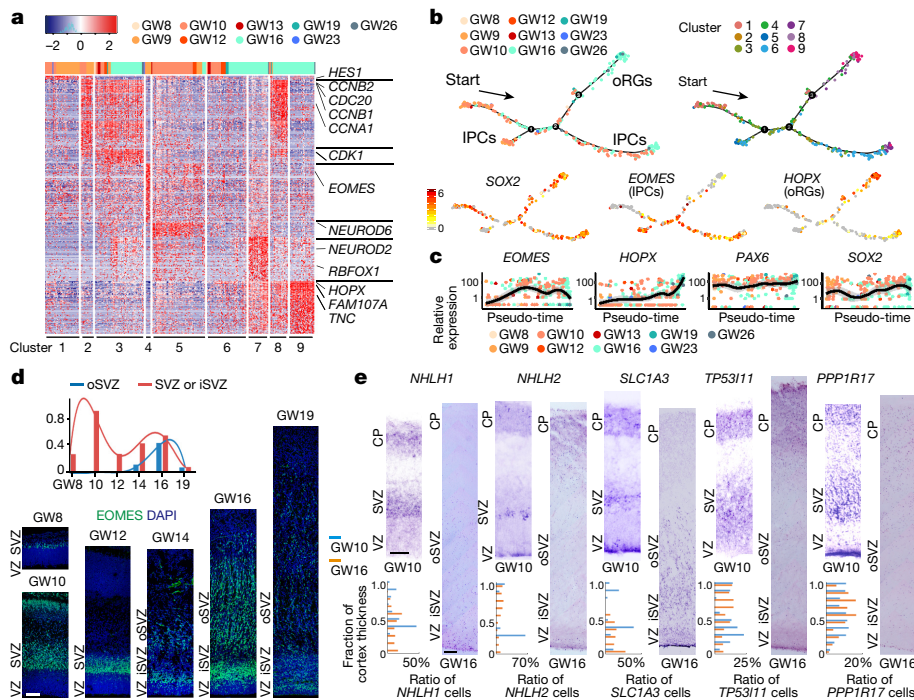


Figure 2 | Molecular signature of neural progenitor cells of the developing human PFC. **a**, Heat map showing gene expression in the NPC subclasses segregated using random forest classification. Top, the distribution of each subclass in gestational weeks. **b**, Cell lineage relationship of NPCs in the PFC. Monocle recovered a branched single-cell trajectory. **c**, Expression of NPC markers, ordered by Monocle analysis in pseudo-time. The shadow represents the confidence interval around the fitted curve. $n = 288$ NPCs. **d**, EOMES⁺ IPCs in the PFC at different gestational times. Bar charts represent the ratio of EOMES⁺ cells in the iSVZ or SVZ (red line) or oSVZ (blue line) in **c**, ventricular zone/ **e**, *In situ* hybridization of IPC-specific genes in the PFC at GW10 and GW16. The cells expressing IPC markers were distributed across ten evenly spaced bins going from the ventricular zone to the cortical plate (CP). The experiments in **c–e** were repeated three times independently with similar results. Scale bars, 100 μ m (**d**, **e**). See Extended Data Figs 6, 7.

and GW16 (Extended Data Fig. 6e). To validate these results, we immunostained cortices of the PFC from GW8 to GW19 (Fig. 2d). We observed early increases in the number of IPCs at GW8, just above the thick ventricular zone. At GW10, the number of IPCs increased rapidly in the expanded SVZ. Thereafter, IPCs were not generated actively until GW16, when they were abundant in the iSVZ and oSVZ. Quantification corroborated the two peaks of IPC proliferation at GW10 and GW16, primarily located in the SVZ and oSVZ, respectively (Fig. 2d). These results suggest that accumulated IPCs at GW10 might originate from the vRG cells, whereas the majority of IPCs, especially those located in the oSVZ, might originate from the oRG cells that are present at GW16. To further characterize the differences between IPCs generated at early and late developmental stages, we compared differentially expressed genes in EOMES⁺ NPCs from GW10 and GW16. IPCs from GW10 and GW16 were enriched for genes involved in stem cell proliferation and neuronal differentiation, respectively (Extended Data Fig. 7a). In addition, we observed some EOMES⁺ cells that expressed early neuronal markers in an intermediate state between cell cycle-active NPCs and cell cycle-inactive EOMES⁺ neurons (Extended Data Fig. 7b). To identify markers for IPCs, we selected genes that were highly expressed in clusters 3, 4 and 5, including *NHLH1*, *NHLH2*, *TP53I11*, *SLC1A3* and *PPP1R17* (Extended Data Fig. 7c, d). We used RNA *in situ* hybridization (ISH) to validate these candidates. *NHLH1*- and *SLC1A3*-expressing cells were primarily localized in the SVZ at GW10, whereas they expanded from the ventricular zone into the oSVZ in GW16 (Fig. 2e). *NHLH2*, *TP53I11* and *PPP1R17*, in addition to expression in the SVZ at GW10 and in the oSVZ at GW16, also exhibited some cortical plate localization (Fig. 2e). These results demonstrate the heterogeneity of IPCs and identify new markers for IPCs that are expressed at different developmental time points.

The two main types of neurons in the brain are excitatory neurons and interneurons^{3,24,25}. Excitatory neurons could be further divided into seven clusters by random forest analysis (Extended Data Fig. 8a), and these were categorized into three groups according to their developmental stage (Fig. 3a–c, Extended Data Fig. 8b). Gene ontology terms enriched in each of the adjacent two periods showed that gene sets related to regulation of cell migration were more enriched at GW13 than at GW12, suggesting that GW13 is a critical time for migration of newly formed neurons. Neurons from GW16 PFC showed increased expression of genes for cell fate commitment; whereas neurons from

GW19 PFC exhibited increased expression of functional genes, such as those involved in calcium import. Genes related to axonogenesis were expressed in late stage PFC (GW19 to GW26); this was followed by expression of genes for synapse formation from GW23 to GW26 (Fig. 3c, Extended Data Fig. 8c), suggesting that the initial formation of neural connections takes place between GW19 and GW26.

To confirm this hypothesis, we examined the electrophysiological properties of neurons in different layers of the PFC at GW23 and GW26. We analysed 53 neurons from three GW23 PFC samples. No action potentials were detected in these neurons (Extended Data Fig. 9a–c). Next, we examined 58 neurons from the cortical plate and subplate of three GW26 PFC biological samples. In the upper layer neurons (SATB2⁺FOXP2[−]), action potentials were only barely initiated, with a low level of outward potassium and inward sodium channel currents (Fig. 3d–f, Extended Data Fig. 9d). Deeper layer neurons (SATB2⁺FOXP2⁺) exhibited stronger action potentials and patterned firing (Fig. 3d–f, Extended Data Fig. 9d). There were no differences between whole-cell membrane Na⁺–K⁺ currents in upper- or deep-layer neurons, but deep-layer neurons exhibited more complex 3D morphology (Fig. 3d, g–i, Extended Data Fig. 9d–f). In addition, both spontaneous EPSCs (excitatory postsynaptic currents) and IPSCs (inhibitory postsynaptic currents) were detected in several deep-layer neurons (Fig. 3j), suggesting that these cells were beginning to receive excitatory and inhibitory presynaptic inputs. Subplate neurons also showed more mature electrophysiology than the cortical plate neurons (Fig. 3d–h, Extended Data Fig. 9d). Consistent with their mature electrophysiological properties, the subplate neurons exhibited more complex 3D dendrite morphology than neurons located in the upper or deep layers (Fig. 3d, i, Extended Data Fig. 9f).

Cortical interneurons originate from the ganglionic eminence and migrate tangentially to the neocortex, where excitatory neurons are generated^{10–12}. However, whether a subgroup of interneurons is generated locally in the PFC remains unclear. Random forest analysis clusters PFC interneurons into eight subgroups (Extended Data Fig. 10a). Groups of cells expressing interneuron progenitor markers, such as *TTF1*, *LHX6* and *DLX1*, were detected at an early timepoint in a pseudo-time alignment by Monocle analysis, and persisted throughout development (Fig. 4a–c). *CALB2*⁺ and *SST*⁺ interneurons appeared early in pseudo-time, and were followed by *CALB1*⁺, *CCK*⁺ and *VIP*⁺ interneurons (Fig. 4 a–c, Extended Data Fig. 10b).

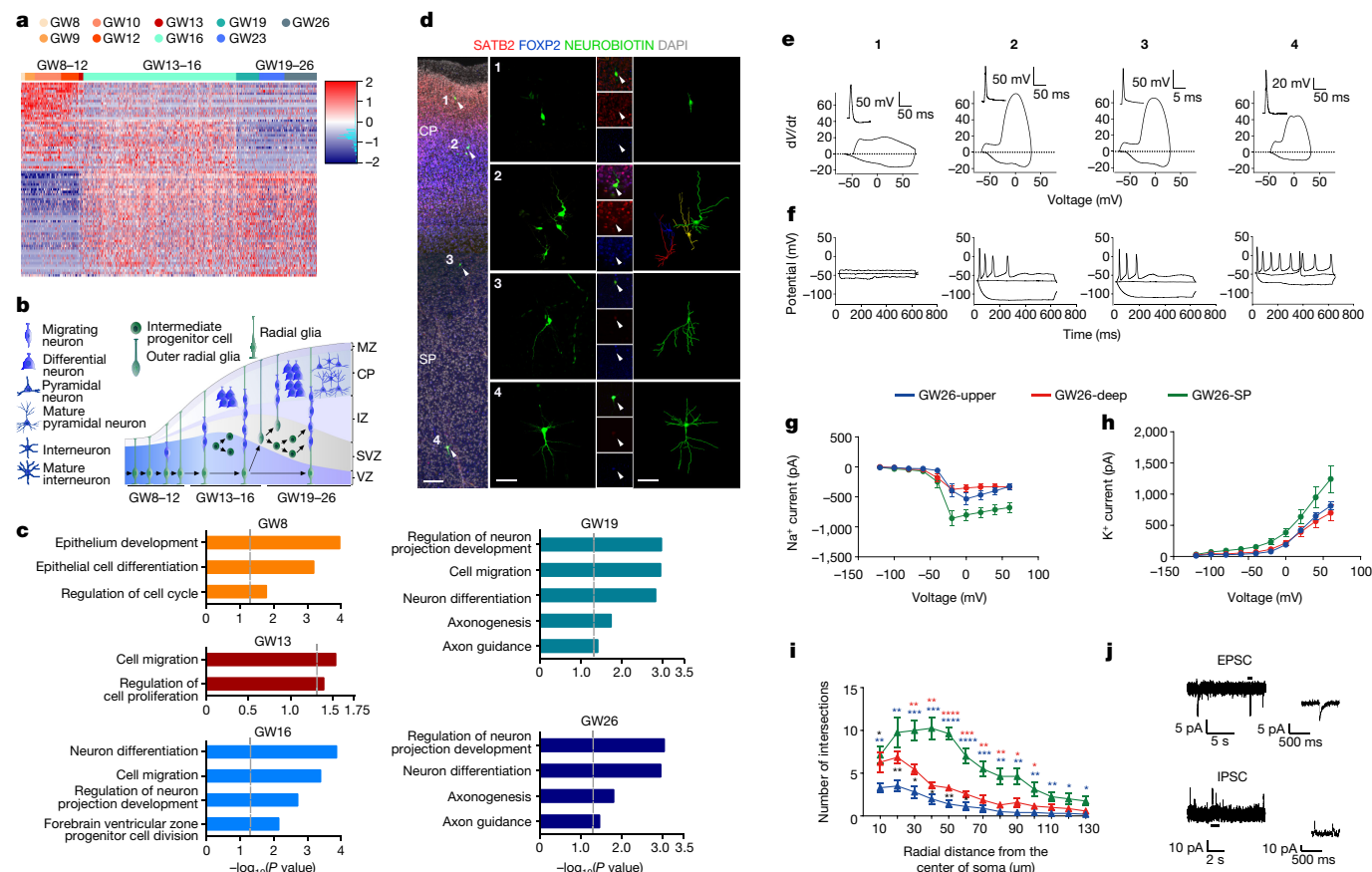


Figure 3 | Dynamics of neurogenesis in the developing human PFC.

a, Heat map shows differentially expressed genes in excitatory neurons from the three developmental stages. **b**, Scheme showing the timing of neurogenesis in the developing PFC. MZ, marginal zone; IZ, intermediate zone. **c**, The enriched gene ontology terms show the cell properties of the PFC at different weeks. $n = 1,057$ neurons. **d**, Immunostaining after whole-cell patch-clamp recordings of PFC neurons at GW26. SP, subplate. Right, 3D reconstructions of traced cells. Scale bar, 50 μm . Representative images are shown from $n = 26$ neurons of three independent replicates. **e**, **f**, Whole-cell patch-clamp recordings including action potential (insets in **e**), phase plot (**e**) and evoked action potential (**f**) of the PFC neurons in **d** at GW26. **g**, **h**, Quantification of

Na^+ current (**g**) and K^+ current (**h**) in neurons from different locations in GW26 PFC. $n = 42$ neurons from three independent replicates. Upper versus deep, $P > 0.05$; upper versus subplate, $P < 0.0001$; deep versus subplate, $P < 0.0001$; two-way analysis of variance (ANOVA). **i**, Quantification of neuron dendrite morphology by Sholl analysis. $n = 25$ from three independent replicates. Asterisks indicate P values for comparisons between neurons from upper and deep layers (black), deep layer and subplate (red), upper layer and subplate (blue): * $P < 0.05$, ** $P < 0.01$, *** $P < 0.001$, **** $P < 0.0001$; two-tailed t -test. **j**, Representative example of whole-cell patch-clamp recordings of spontaneous EPSC and IPSC of deeper layer neurons. Data in **g**–**i** are mean \pm s.e.m. See Extended Data Figs 8, 9.

No parvalbumin (PV)⁺ interneurons were detected, suggesting that this neuron subtype might develop after GW26 (Fig. 4c). To assess whether interneurons are generated locally in the developing human PFC, we examined the presence and localization of interneuron progenitors by ISH and immunofluorescence. We observed a few TTF1⁺ cells in the anterior dorsal PFC above the classical neural progenitor zone at GW10 (Fig. 4d, Extended Data Fig. 10c). By contrast, TTF1⁺ cells were enriched in the medial ganglionic eminence (MGE), but not in the cortex close to the MGE (Fig. 4d). There were more TTF1⁺ cells in the dorsal PFC at GW17, which may originate from the ganglionic eminence (Fig. 4d). Furthermore, cell cycle analysis revealed that very few *TTF1*, *LHX6*, *DLX1* or *DLX2*-expressing cells in the developing PFC were actively progressing through the cell cycle (Extended Data Fig. 10d). Notably, *SST* mRNA could be detected in the PFC as early as GW7 by PCR with reverse transcription and in GW8 by RNA sequencing (RNA-seq) analysis (Fig. 4c, Extended Data Fig. 10e). This early appearance of *SST*-expressing cells suggests that they are unlikely to have migrated from the ganglionic eminence. However, using ISH, we also detected *SST* mRNA in the ventricular zone and subventricular zone at GW10, although *SST* protein was undetectable at this stage (Extended Data Fig. 10f, g). Together, our data indicate that although cells expressing interneuron progenitor markers were

detected in the dorsal PFC at an early stage, they may not have entered the cell cycle.

Since excitatory neurons and interneurons build up circuits cooperatively, we next investigated whether the development of these two types of neurons was synchronized. Excitatory neuron development peaked at GW16, whereas interneuron development peaked at GW26. We therefore selected neurons at GW16 and GW26 for pathway analysis by gene-set enrichment analysis (GSEA). Axon guidance signals, including axon attraction, repulsion and outgrowth, were more active in excitatory neurons than in interneurons at GW16; this was reversed at GW26 (Fig. 4e, f). The neurotrophin signalling pathway, which has roles in neuronal differentiation and prevention of cell death, exhibited a similar pattern (Extended Data Fig. 10h). Additionally, gene enrichment analysis showed that Notch signals were more involved in regulating biological activities of NPCs than those of neurons (Extended Data Fig. 10i). Together, these results suggest that excitatory neurons mature earlier than interneurons, and that this maturation process is regulated by extrinsic and intrinsic signals.

Our single-cell-resolution data illustrate the complex diversity of cell types in the developing human PFC that underlies the sophisticated cognitive function of humans. Notably, we found new IPC markers that enabled us to identify diverse neuron subtypes that are

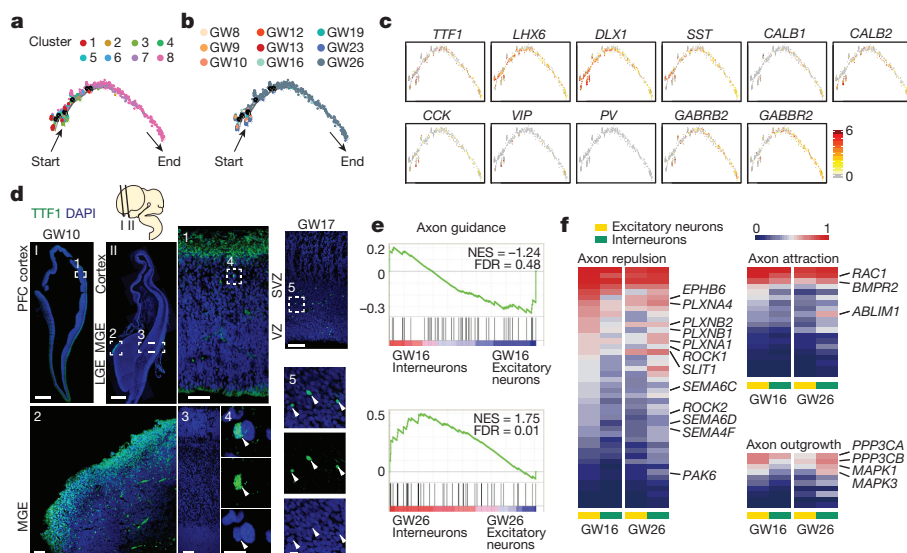


Figure 4 | Interneuron development and signalling pathways that regulate neuronal maturation. **a, b**, Single-cell trajectories of interneurons in the PFC by Monocle analysis. **c**, Expression of GABA receptors and genes known to be expressed in interneuron progenitors and interneuron subtypes are mapped to the single-cell trajectory plot. **d**, Immunofluorescence staining of TTF1 in the PFC (section I) and MGE-LGE (section II) at GW10 and GW17. LGE, Lateral ganglionic eminence. Scale bars, 500 μ m (I, II), 100 μ m (top right (GW17), 3), 50 μ m (1, 2) and 10 μ m (4, 5). $n = 4$ independent replicates per gestational week. **e**, GSEA of axon guidance signalling pathway in excitatory neurons and interneurons at GW16 and GW26. $n = 663$ excitatory neurons; $n = 485$ interneurons. NES, normalized enrichment score; FDR, false discovery rate. **f**, Mean expression of genes involved in the axon guidance signalling pathway in excitatory neurons and interneurons at GW16 and GW26. See Extended Data Fig. 10.

likely to contribute to the cellular basis of elaborate circuit formation in humans. Some neurological disorders and social cognition deficits, such as autism spectrum disorder and schizophrenia, have been linked to an imbalance of excitatory and inhibitory neurons (E/I ratio) in the PFC^{26–28}. Our data indicate that interneurons appear and mature later than excitatory neurons, and that this pattern is intrinsically regulated. Thus, transcriptome profiling of thousands of single cells in the developing PFC is a powerful tool for investigating the mechanisms behind neurological diseases and exploring potential therapies.

Online Content Methods, along with any additional Extended Data display items and Source Data, are available in the online version of the paper; references unique to these sections appear only in the online paper.

Received 15 September 2017; accepted 7 February 2018.

Published online 14 March 2018.

- Roth, G. & Dicke, U. Evolution of the brain and intelligence in primates. *Prog. Brain Res.* **195**, 413–430 (2012).
- O’Rahilly, R. & Müller, F. Significant features in the early prenatal development of the human brain. *Ann. Anat.* **190**, 105–118 (2008).
- Rakic, P. Evolution of the neocortex: a perspective from developmental biology. *Nat. Rev. Neurosci.* **10**, 724–735 (2009).
- Masserdotti, G., Gascón, S. & Götz, M. Direct neuronal reprogramming: learning from and for development. *Development* **143**, 2494–2510 (2016).
- Satija, R., Farrell, J. A., Gennert, D., Schier, A. F. & Regev, A. Spatial reconstruction of single-cell gene expression data. *Nat. Biotechnol.* **33**, 495–502 (2015).
- Lake, B. B. *et al.* Neuronal subtypes and diversity revealed by single-nucleus RNA sequencing of the human brain. *Science* **352**, 1586–1590 (2016).
- Trapnell, C. *et al.* The dynamics and regulators of cell fate decisions are revealed by pseudotemporal ordering of single cells. *Nat. Biotechnol.* **32**, 381–386 (2014).
- Qiu, X. *et al.* Reversed graph embedding resolves complex single-cell trajectories. *Nat. Methods* **14**, 979–982 (2017).
- Qiu, X. *et al.* Single-cell mRNA quantification and differential analysis with Census. *Nat. Methods* **14**, 309–315 (2017).
- Ma, T. *et al.* Subcortical origins of human and monkey neocortical interneurons. *Nat. Neurosci.* **16**, 1588–1597 (2013).
- Sandberg, M. *et al.* Transcriptional networks controlled by NKX2-1 in the development of forebrain GABAergic neurons. *Neuron* **91**, 1260–1275 (2016).
- Radonjić, N. V. *et al.* Diversity of cortical interneurons in primates: the role of the dorsal proliferative niche. *Cell Rep.* **9**, 2139–2151 (2014).
- Kierdorf, K. *et al.* Microglia emerge from erythromyeloid precursors via Pu.1- and Irf8-dependent pathways. *Nat. Neurosci.* **16**, 273–280 (2013).
- Prinz, M. & Priller, J. Microglia and brain macrophages in the molecular age: from origin to neuropsychiatric disease. *Nat. Rev. Neurosci.* **15**, 300–312 (2014).
- Hong, S. *et al.* Complement and microglia mediate early synapse loss in Alzheimer mouse models. *Science* **352**, 712–716 (2016).
- Vasek, M. J. *et al.* A complement-microglial axis drives synapse loss during virus-induced memory impairment. *Nature* **534**, 538–543 (2016).
- Pollen, A. A. *et al.* Molecular identity of human outer radial glia during cortical development. *Cell* **163**, 55–67 (2015).
- Lewitus, E., Kelava, I. & Huttner, W. B. Conical expansion of the outer subventricular zone and the role of neocortical folding in evolution and development. *Front. Hum. Neurosci.* **7**, 424 (2013).

- Noctor, S. C., Martínez-Cerdeño, V., Ivic, L. & Kriegstein, A. R. Cortical neurons arise in symmetric and asymmetric division zones and migrate through specific phases. *Nat. Neurosci.* **7**, 136–144 (2004).
- Homem, C. C., Repic, M. & Knoblich, J. A. Proliferation control in neural stem and progenitor cells. *Nat. Rev. Neurosci.* **16**, 647–659 (2015).
- Kowalczyk, T. *et al.* Intermediate neuronal progenitors (basal progenitors) produce pyramidal-projection neurons for all layers of cerebral cortex. *Cereb. Cortex* **19**, 2439–2450 (2009).
- Noctor, S. C., Martínez-Cerdeño, V. & Kriegstein, A. R. Contribution of intermediate progenitor cells to cortical histogenesis. *Arch. Neurol.* **64**, 639–642 (2007).
- Vasistha, N. A. *et al.* Cortical and clonal contribution of Tbr2 expressing progenitors in the developing mouse brain. *Cereb. Cortex* **25**, 3290–3302 (2015).
- Lui, J. H., Hansen, D. V. & Kriegstein, A. R. Development and evolution of the human neocortex. *Cell* **146**, 18–36 (2011).
- Bandler, R. C., Mayer, C. & Fishell, G. Cortical interneuron specification: the juncture of genes, time and geometry. *Curr. Opin. Neurobiol.* **42**, 17–24 (2017).
- Shi, G. X., Han, J. & Andres, D. A. Rin GTPase couples nerve growth factor signaling to p38 and b-Raf/ERK pathways to promote neuronal differentiation. *J. Biol. Chem.* **280**, 37599–37609 (2005).
- Canitano, R. & Pallagrosi, M. Autism spectrum disorders and schizophrenia spectrum disorders: excitation/inhibition imbalance and developmental trajectories. *Front. Psychiatry* **8**, 69 (2017).
- Bicks, L. K., Koike, H., Akbarian, S. & Morishita, H. Prefrontal cortex and social cognition in mouse and man. *Front. Psychol.* **6**, 1805 (2015).

Supplementary Information is available in the online version of the paper.

Acknowledgements We thank members of the Wang and Tang laboratories for discussions. This work was supported by National Basic Research Program of China (2014CB964600), the Strategic Priority Research Program of the Chinese Academy of Sciences (XDA16020601), the National Natural Science Foundation of China (NSFC) (91732301, 31371100, 31771140), National Key Research and Development Program of China (2017YFA0103303, 2017YFA0102601), Shanghai Brain-Intelligence Project from STCSM (16JC1420500), Newton Advanced Fellowship (NA140246) to X.W. and Youth Innovation Promotion Association CAS to Q.W.

Author Contributions Q.W., X.F., J.Q., F.T. and X.W. conceived the project, designed the experiments and wrote the manuscript. S.Zho. and X.F. performed RNA-seq. S.Zha. and J.D. analysed the data. J.Z., L.L., L.S., H.Z., L.Y. and X.X. prepared the samples. S.Zho. and Q.W. performed immunofluorescence, *in situ* hybridization and imaging. L.S. and N.P. performed the electrophysiology experiments. All authors edited and proofread the manuscript.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing interests. Readers are welcome to comment on the online version of the paper. Publisher’s note: Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations. Correspondence and requests for materials should be addressed to X.W. (xiaoqunwang@ibp.ac.cn), J.Q. (jie.qiao@263.net), J.Z. (drzhangji@outlook.com) or F.T. (tangfuchou@pku.edu.cn).

Reviewer Information Nature thanks H. Song and the other anonymous reviewer(s) for their contribution to the peer review of this work.

METHODS

Ethics statement. The human embryo collection and research analysis was approved by the Reproductive Study Ethics Committee of Peking University Third Hospital (2012SZ-013 and 2017SZ-043) and Beijing Anzhen Hospital (2014012x). The informed consent was designed as recommended by the ISSCR guidelines for fetal tissue donation. Informed consent for fetal tissue procurement and research was obtained from the patient after her decision to legally terminate her pregnancy but before the abortive procedure. Fetal cortical tissue samples were collected after the donor patients signed an informed consent document that was in strict observance of the legal and institutional ethical regulations for elective pregnancy termination specimens at Peking University Third Hospital and Beijing Anzhen Hospital, Capital Medical University. All samples used in these studies had not been involved in any other procedures. All the protocols were in compliance with the 'Interim Measures for the Administration of Human Genetic Resources' administered by the Chinese Ministry of Health.

Tissue sample collection and dissection. Fetal brains were collected in ice-cold artificial cerebrospinal fluid containing 125.0 mM NaCl, 26.0 mM NaHCO₃, 2.5 mM KCl, 2.0 mM CaCl₂, 1.0 mM MgCl₂, 1.25 mM NaH₂PO₄ at a pH of 7.4 when oxygenated (95% O₂ and 5% CO₂). The prefrontal cortex was dissected and placed in Hibernate E medium (Invitrogen, A1247601). The PFC tissue was first digested in brain tissue digestion medium 1 (BTDM1, 2 mg/ml collagenase IV (Gibco, 17104-019) and 10 U/μl DNase I (NEB, M0303L) in Hibernate E medium). After brief pipetting to roughly separate the tissue into small pieces, brain tissue was then digested in digestion medium 2 (BTDM2, 1 mg/ml papain (Sigma, P4762) and 10 U/μl DNase I in Hibernate E medium). The tissue was vortexed and kept at 37°C for 5 min. The cell suspension was pipetted further to disperse it into single cells and centrifuged at 300g for 5 min to obtain a cell pellet. The digestion medium was carefully removed and the cell pellet was resuspended in 500 μl Hibernate E medium and kept on ice.

Single cells were picked into cell lysis buffer by mouth pipette and reverse transcription was performed as described for Smartseq2^{29,30}, except that we used barcoded reverse transcription primers instead of the oligdT30VN primer. Amplification was performed using the Smartseq2 protocol with minor modifications. The PCR products with different barcodes were pooled together for purification and library construction³¹. The libraries were processed on the Illumina platform for sequencing of 150 bp pair-end reads.

Processing of raw single-cell RNA-seq data. The read2 data of pooled cells were split into single-cell data using the barcode sequences contained in the first 8 bp. The next 8 bp were recorded as unique molecular identifiers (UMIs). The read1 data were then sorted by the header of the corresponding read2 data and added to the UMI information. The poly(A) tail sequence and template switch oligo (TSO) sequences were then trimmed from the read1 sequences. Quality control was performed by removing reads with adapter contaminants (length < 37 bp) and low-quality bases ($N > 10\%$). Subsequently, TopHat³² (version 2.0.12) was used to align clean reads to the hg19 human transcriptome downloaded from UCSC³³. Uniquely aligned reads were counted using the 'htseq-count' tool of HTSeq³⁴. Transcripts for the same genes that shared the same UMI sequence were merged, thus the number of different UMIs for each gene was considered as the transcript count of that gene without PCR bias for each sequenced cell.

The transcript counts of each cell were normalized to transcript per million (TPM), where TPM is the transcript count of each gene divided by the sum of transcript counts of that cell, multiplied by one million. TPM values were then normalized by $\log((\text{TPM}/10) + 1)$ for subsequent analysis. TPM was divided by 10 because the sequencing depth of 64% of single-cell samples contained less than 1,000,000 reads. Overall 2,394 individual cells were collected for single-cell cDNA amplification and 2,309 cells passed the quality control criteria. On average, there were 1.1 million mapped reads and 2,654 detected genes for each cell.

Identification of cell types and subtypes by nonlinear dimensional reduction and random forests. The Seurat⁵ package (v.1.2.1) implemented in R was applied to identify major cell types among 2,394 single cells from the PFC. Only cells that expressed more than 1,000 genes were considered, and only genes with normalized expression level greater than 1 that were expressed in at least three single cells were included, leaving 17,854 genes across 2,333 samples for clustering analysis. After initial clustering, two clusters enriched in haemoglobin genes and microglia-specific genes (*PTPRC*, *CSF1R*, *AIF1*) were removed before the second clustering analysis. The haemoglobin genes *HBA1*, *HBA2*, *HBB*, *HBD*, *HBE1*, *HBG1*, *HBG2*, *HBM*, *HBQ1* and *HBZ*, and immune cells found in the second clustering analysis were also excluded from subsequent analyses. Following the second clustering analysis, 2,309 cells excluding immune cells and cells enriched in haemoglobin genes were mapped by *t*-SNE (Fig. 1a). *PAX6*, *NEUROD2*, *GAD1*, *PDGFRA*, *AQP4* and *PTPRC* were used as markers to identify the major cell types in the brain: neural progenitor cells (NPCs), excitatory neurons, interneurons, oligodendrocyte progenitor cells (OPCs), astrocytes and microglia, respectively.

Random forests⁶ was used to identify subtypes of the major cell types in human embryonic PFC. We used about 1,000 genes, identified as over-dispersed genes by 'mean.var.plot' of Seurat, as feature genes for selection by random forests. The number of folds in the cross-validation of random forest analysis was varied between 5 and 10.

Identification of differentially expressed genes among clusters. Genes that were differentially expressed in each cluster were identified using the Seurat function 'find_all_markers' and tested by 'roc' and DESeq2²⁵. The ROC test returns the 'classification power' for any individual gene (ranging from 0 (random) to 1 (perfect)); positive 'avg_diff' means upregulation in the corresponding clusters, whereas negative 'avg_diff' means downregulation. DESeq2 returns 'log2Fold-Change' (higher absolute value means higher fold change of corresponding clusters) and *P* value of each tested gene. We used DAVID^{36,37} (<https://david.ncifcrf.gov/home.jsp>) and Metascape³⁸ (<http://metascape.org>) to perform biological process enrichment analysis with the differentially expressed genes of each cluster.

Constructing single cell trajectories in the PFC. The Monocle⁷⁻⁹ package was used to analyse single cell trajectories in order to discover developmental transitions. We used differentially expressed genes identified by Seurat to sort cells in pseudo-time order. The actual gestational time of each cell informed us of the start point of the pseudo-time in the first round of 'orderCells'. We then set this state as the root_state argument and called 'orderCells' again. 'DDRTree' was applied to reduce dimensions and the visualization functions 'plot_cell_trajectory' or 'plot_complex_cell_trajectory' were used to plot the minimum spanning tree on cells.

Pathways analysis. GSEA³⁹ was applied to identify a priori-defined gene sets that show statistically significant differences between two given clusters (<http://www.broadinstitute.org/gsea/index.jsp>). We used the expression file containing about 5,000 variable genes as input, and implied gene sets of KEGG⁴⁰ pathways and Gene Ontology⁴¹ (<http://www.geneontology.org/>), which were collected in Molecular Signatures Database (MSigDB)^{39,42}.

We calculated the mean expression of 'axon guidance pathway' genes in excitatory and inhibitory neurons respectively from GW16 and GW26 PFC. The mean expression values of both types of neurons at the same gestational time were compared and tested by *t*-test for significance. Significantly differently expressed genes were plotted as heat maps.

The average gene expression level for each stage (early, middle and late stages) in the Notch pathway was plotted in the heat map and the ratio of expressed genes to all the genes in Notch pathway was plotted as a histogram at the top of the heat map.

Cell-cycle analysis. We applied a cell-cycle-related gene set with 46 genes for G1/S phase and 54 genes for G2/M phase of the cell cycle. We added three genes, *CDK4*, *CDK2* and *CDK6*, based on the core set of cell-cycle-related genes from Tirosh *et al.*^{43,44}. We defined G1/S and G2/M states of each cell by comparing the average expression of the two gene sets.

Immunofluorescent staining. Tissue samples were fixed overnight in 4% paraformaldehyde, cryoprotected in 30% sucrose and embedded in optimal cutting temperature medium (Thermo Scientific). Thin 40-μm cryosections were collected on superfrost slides (VWR) using a Leica CM3050S cryostat. For immunohistochemistry, heat-induced antigen retrieval was performed in 10 mM sodium citrate buffer, pH 6. Primary antibodies: rabbit anti-FOXP2 (1:500, Abcam ab16046), mouse anti-SATB2 (1:250, Abcam ab51502), mouse anti-CD45 (1:100, Abcam ab8216), goat anti-AIF1 (1:200, Abcam ab5076), goat anti-SOX2 (1:250, Santa Cruz sc-17320), rabbit anti-PAX6 (1:500, BioLegend 901301), rabbit anti-SFRP1 (1:500, Abcam ab126613), chicken anti-EOMES (1:500, Millipore AB15894), mouse anti-RBFOX1 (1:500, Abcam ab183348), rabbit anti-TTF1 (1:300, Abcam ab86023), Neurobiotin tracer (1:1,000, Vectorlabs SP-1120-50), rabbit anti-NEUROD2 (1:500, Abcam ab104430), mouse anti-NEUROD1 (1:100, Abcam ab60704), rabbit anti-HMGA2 (1:100, Abcam ab207301), rabbit anti-HOPX (1:1,000, Santa Cruz sc-30216), mouse anti-Ki-67 (1:100, BD 550609). Primary antibodies were diluted in blocking buffer containing 10% donkey serum, 0.5% Triton X-100 and 0.2% gelatin in PBS at pH 7.4. Binding was visualized using an appropriate Alexa Fluor 488, Alexa Fluor 594 or Alexa Fluor 647 fluorophore-conjugated secondary antibody (Life Technologies). Cell nuclei were stained using DAPI (Life Technologies). Images were collected using an Olympus FV1000 confocal microscope.

In situ hybridization. The *in situ* hybridization protocol has been described previously⁴⁵. In brief, probes complementary to target human mRNA used for RNA *in situ* hybridization were cloned from primary human fetal cortical cDNA samples, reverse-transcribed using PrimeScript II 1st Strand cDNA Synthesis Kit (Takara) with oligo dT primers; or from RNA samples isolated from GW7 or GW17 human cortex using SV Total RNA Isolation System (Promega). Specific genes were amplified using the following primers: *NHLH1* forward, CTTCCGCGAGCTGCGCAAG; *NHLH1* reverse, ATTCCCAGCTACTGAGATTC; *NHLH2* forward, GAGCTCCGCAATTGCTGC; *NHLH2* reverse, GCTCAGATCCTCCACAGT;

TP53I11 forward, CGTGTGCTGTGATCTCCTGG; *TP53I11* reverse, GAACTCAGGCTGGTTGCAG; *SLC1A3* forward, CTAGGCCTCAGTGCCTCATCT; *SLC1A3* reverse, GCGTCTTTGACTGGATATTCCT; *PPP1R17* forward, TGTGACATTGCTCAGGGACG; *PPP1R17* reverse, CTCCTAGCTAA CACCACACTC; *SST* forward, CCTAGAGTTTGACCAGCCACTC; *SST* reverse, AGTTTCTAATGCAAGGGTCTCGP; *TTFI* forward, CGACTTTTC TTAACAACCTGGC; *TTFI* reverse, AAAAGACTGACGCCGCAAATAC. Primers specific for target genes of interest were designed using Primer3 and amplified by PCR using Q5 High-Fidelity DNA Polymerase (NEB). PCR products of the predicted band sizes were extracted from gels and ligated into the Hieff Clone Plus One Step Cloning Kit (Yeaston). Ligation products were transfected into Trans5 α Chemically Competent *Escherichia coli* (Transgene). Cloned sequences were confirmed by sequencing. Digoxigenin-labelled RNA probes for *in situ* hybridization were generated by linearizing the pSPT18 Vector and transcribing the probe *in vitro* using T7 or SP6 RNA Polymerase (Roche) in the presence of DIG-RNA Labelling Mix (Roche). Fetal brain sections (30 μ m thick) were hybridized with RNA probes, at a final concentration of 500 ng/ml, overnight at 64.5 °C in hybridization solution (50% formamide, 10% dextran sulphate, 0.2% tRNA (Invitrogen), 1 \times Denhardt's solution (Sigma), 1 \times salt solution (containing 0.2 M NaCl, 0.01 M Tris, 5 mM NaH₂PO₄, 5 mM Na₂HPO₄, 5 mM EDTA pH 7.5)). After the sections were washed, alkaline phosphatase-coupled anti-digoxigenin Fab fragments (Roche) were applied. For visualization of the labelled cRNAs, the sections were incubated in the dark in NBT–BCIP solution (Roche). Images were taken with a Leica SCN400 (Leica Microsystems).

Electrophysiological recording. Coronal slices containing PFC (500 μ m) from GW23 or GW26 embryos were prepared using a vibratome (VT1200S, Leica) in oxygenated (95% O₂ and 5% CO₂) ice-cold sucrose-based artificial cerebrospinal fluid (s-ACSF, 234 mM sucrose, 2.5 mM KCl, 26 mM NaHCO₃, 1.25 mM NaH₂PO₄, 11 mM D-glucose, 0.5 mM CaCl₂ and 10 mM MgSO₄). The slices were kept in an incubating chamber filled with oxygenated ACSF (126 mM NaCl, 3 mM KCl, 1.2 mM NaH₂PO₄, 2.4 mM CaCl₂, 1.3 mM MgSO₄, 26 mM NaHCO₃, 10 mM D-glucose) at 34 °C for 30 min. After a recovery period of at least 60 min at room temperature, an individual slice was transferred to a recording chamber and was continuously superfused with oxygenated ACSF (4 ml/min) at room temperature.

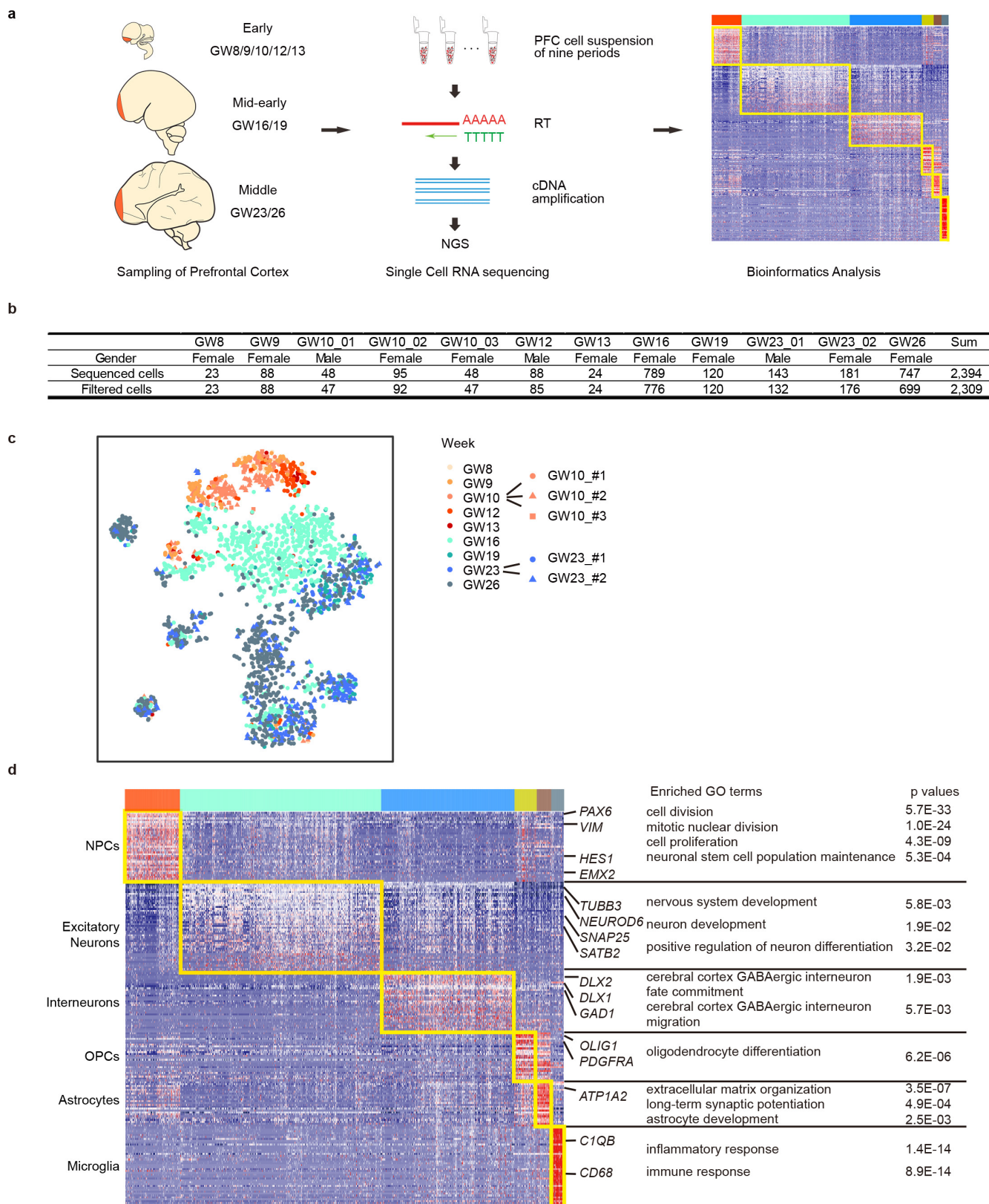
Whole-cell patch-clamp recordings were performed with a patch pipette solution containing (in mM) 140 K-gluconate, 5 KCl, 0.2 EGTA, 2 MgCl₂ and 10 HEPES. The electrode was filled with ~0.5% of neurobiotin and 0.1% Lucifer yellow, which was diffused into the recording neuron under the whole-cell recording for further confirmation staining. Electrode resistance ranged from 10 to 13 M Ω . Tight seals (>1 G Ω) were obtained on cell bodies before rupturing the membrane with negative pressure.

The whole-cell recording was conducted with an MultiClamp 700B amplifier (Axon Instruments). Currents were typically digitized at 100 kHz, and macroscopic records were filtered at 2 kHz. Electrophysiological data were analysed with ClampFit software (version 10.0; Axon Instruments) and GraphPad Prism (v.6.0; GraphPad Software). A single action potential was evoked by injecting short

(2–4 ms) depolarizing current pulses (in the current clamp). Data were accepted for analysis only in cases in which the recorded neurons exhibited a resting potential <–10 mV throughout the experiment. Spiking properties were calculated based on the response to a 500-ms current pulse at –60 pA, 0 pA and 60 pA, respectively. To detect spontaneous glutamate receptor-mediated excitatory postsynaptic currents (sEPSCs) or inhibitory postsynaptic currents (sIPSCs), the cells were monitored at a holding potential of –70 mV or 0 mV, respectively.

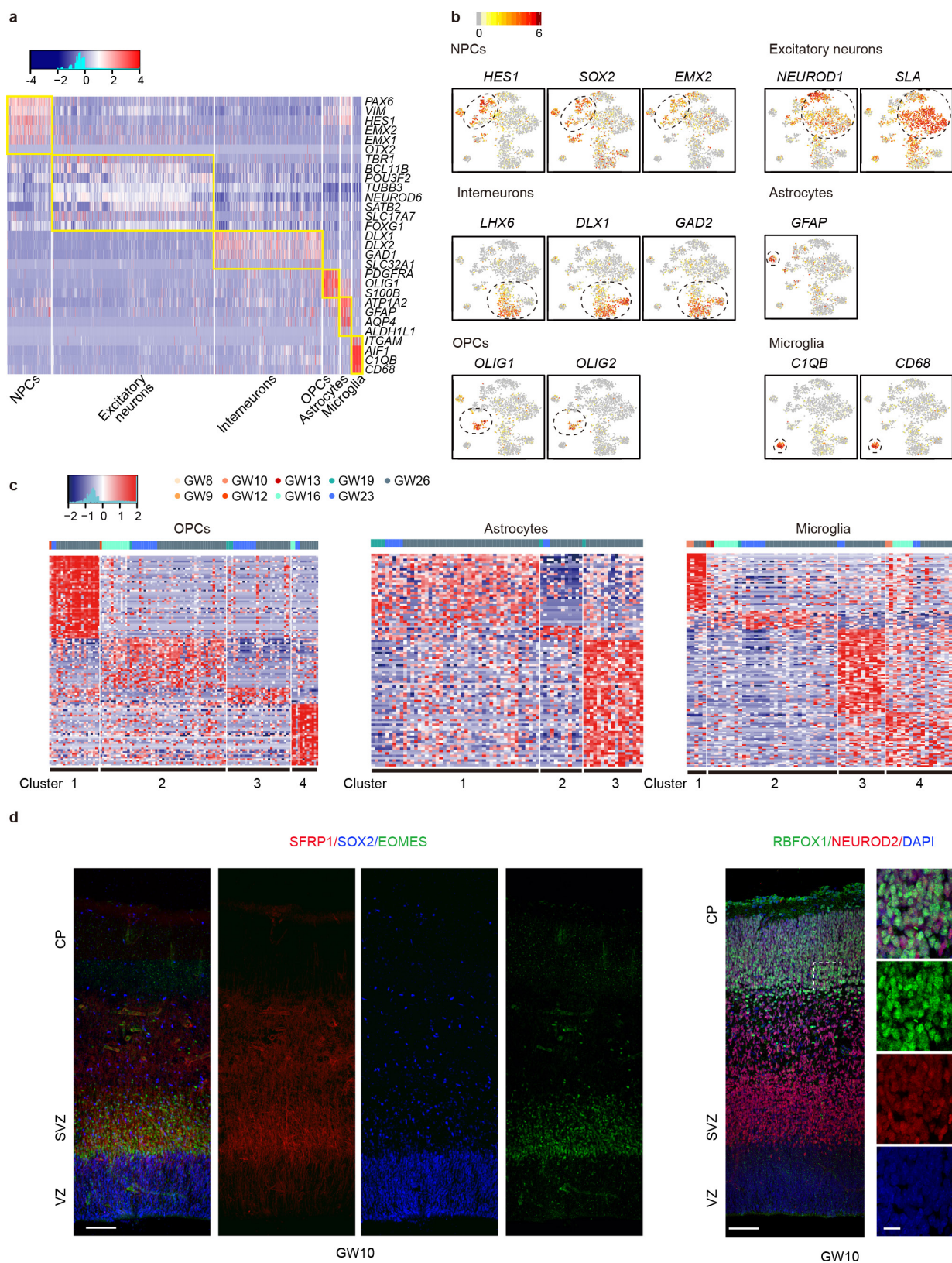
Data availability. The single-cell RNA-seq data used in this study have been deposited in the Gene Expression Omnibus (GEO) under the accession number GSE104276. Raw image files used in the figures that support the findings of this study are available from the corresponding authors upon reasonable request.

29. Picelli, S. *et al.* Full-length RNA-seq from single cells using Smart-seq2. *Nat. Protoc.* **9**, 171–181 (2014).
30. Guo, F. *et al.* The transcriptome and DNA methylome landscapes of human primordial germ cells. *Cell* **161**, 1437–1452 (2015).
31. Li, L. *et al.* Single-cell RNA-seq analysis maps development of human germline cells and gonadal niche interactions. *Cell Stem Cell* **20**, 891–892 (2017).
32. Trapnell, C., Pachter, L. & Salzberg, S. L. TopHat: discovering splice junctions with RNA-seq. *Bioinformatics* **25**, 1105–1111 (2009).
33. Hinrichs, A. S. *et al.* The UCSC Genome Browser Database: update 2006. *Nucleic Acids Res.* **37**, D590–D598 (2006).
34. Anders, S., Pyl, P. T. & Huber, W. HTSeq—a Python framework to work with high-throughput sequencing data. *Bioinformatics* **31**, 166–169 (2015).
35. Love, M. I., Huber, W. & Anders, S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol.* **15**, 550 (2014).
36. Huang, W., Sherman, B. T. & Lempicki, R. A. Bioinformatics enrichment tools: paths toward the comprehensive functional analysis of large gene lists. *Nucleic Acids Res.* **37**, 1–13 (2009).
37. Huang, W., Sherman, B. T. & Lempicki, R. A. Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. *Nat. Protoc.* **4**, 44–57 (2009).
38. Tripathi, S. *et al.* Meta- and orthogonal integration of influenza 'omics' data defines a role for UBR4 in virus budding. *Cell Host Microbe* **18**, 723–735 (2015).
39. Subramanian, A. *et al.* Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. *Proc. Natl Acad. Sci. USA* **102**, 15545–15550 (2005).
40. Ogata, H. *et al.* KEGG: Kyoto encyclopedia of genes and genomes. *Nucleic Acids Res.* **27**, 29–34 (1999).
41. The Gene Ontology Consortium. Gene ontology: tool for the unification of biology. *Nat. Genet.* **25**, 25–29 (2000).
42. Liberzon, A. *et al.* Molecular signatures database (MSigDB) 3.0. *Bioinformatics* **27**, 1739–1740 (2011).
43. Tirosh, I. *et al.* Dissecting the multicellular ecosystem of metastatic melanoma by single-cell RNA-seq. *Science* **352**, 189–196 (2016).
44. Macosko, E. Z. *et al.* Highly parallel genome-wide expression profiling of individual cells using nanoliter droplets. *Cell* **161**, 1202–1214 (2015).
45. Palop, J. J., Roberson, E. D. & Cobos, I. Step-by-step *in situ* hybridization method for localizing gene expression changes in the brain. *Methods Mol. Biol.* **670**, 207–230 (2011).



Extended Data Figure 1 | Single-cell RNA-seq information and molecular diversity of single cells. **a**, Experimental workflow for single-cell RNA-seq of human developing PFC. **b**, Table summarizing PFC sampling. **c**, *t*-SNE plots of cells in the PFC. $n = 3$ (GW10) and $n = 2$ (GW23) independent biological samples. No obvious differences in distribution were observed among the different batches at the same

development stages. Each colour represents the gestational week, and the colour contours correspond to the cell types. Expression of known markers is shown using the same layout (grey, no expression; yellow-red, relative expression). **d**, Heat map shows blocks of genes enriched in each cell type. Right, specific genes related to each type are highlighted with enriched gene ontology terms. $n = 2,309$ cells.

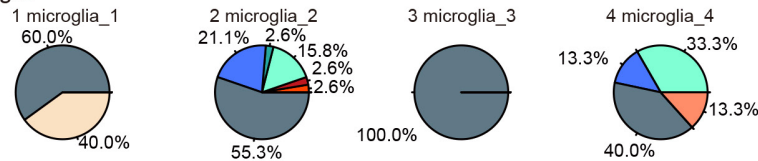


Extended Data Figure 2 | Molecular diversity of subgroups of cells.

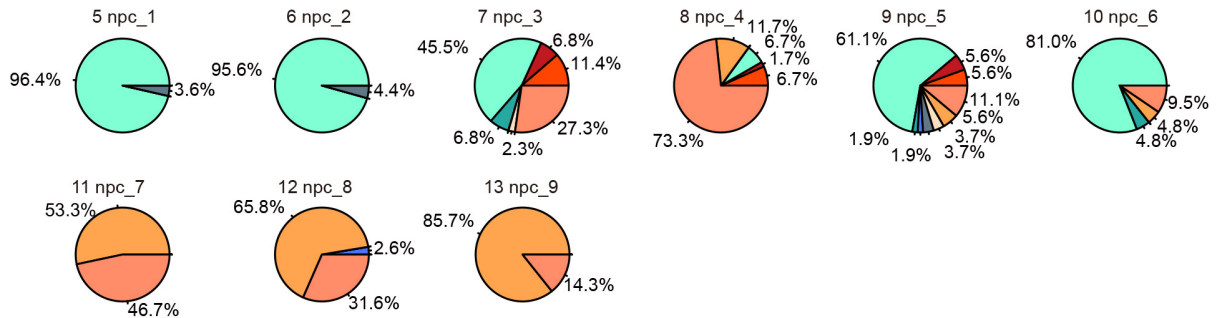
a, Heat map showing the expression level and identity of genes in six major cell types in the PFC. Expression of genes known to be expressed in each cell type is shown to the right of each heat map panel. **b**, Expression of known markers is shown using the same layout as in Fig. 1a (grey, no expression; yellow–red, relative expression). **c**, Heat maps show the

subclasses of OPCs, astrocytes and microglia. The genes were organized into clusters. Top chart, gestational weeks. **d**, Immunostaining for new markers of NPCs (SFRP1) and excitatory neurons (RBFOX1) at GW10 in the PFC. Scale bars, 100 μ m (left), 100 μ m (middle) and 10 μ m (right). $n = 3$ independent replicates per gestational week.

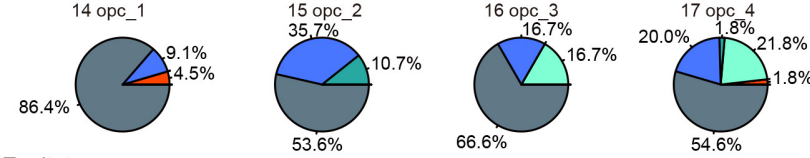
Microglia



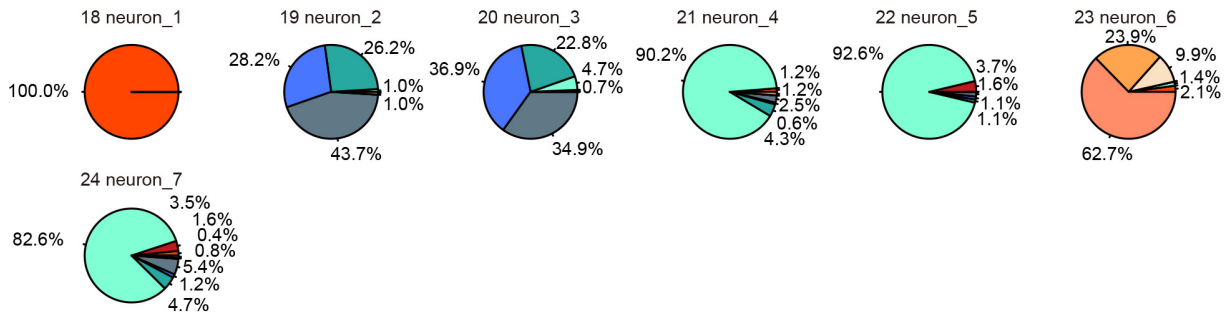
NPCs



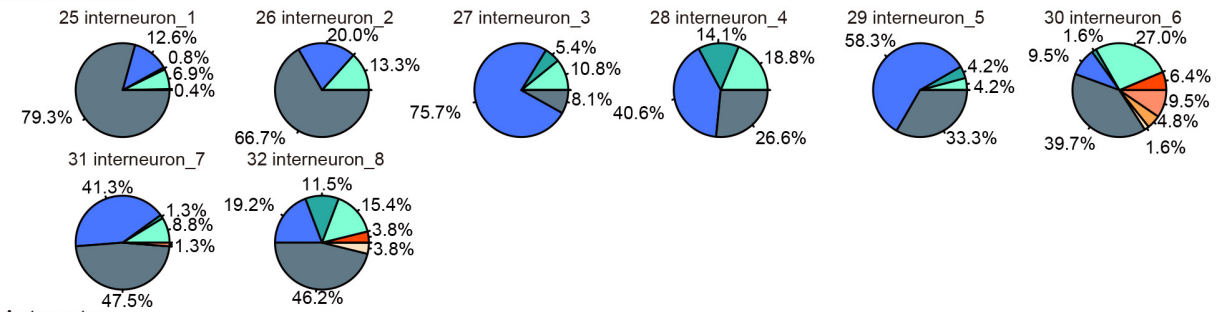
OPCs



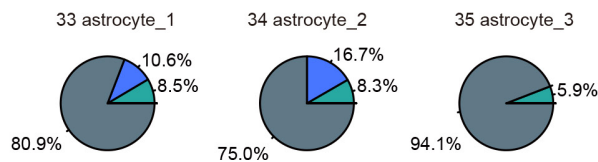
Excitatory neurons



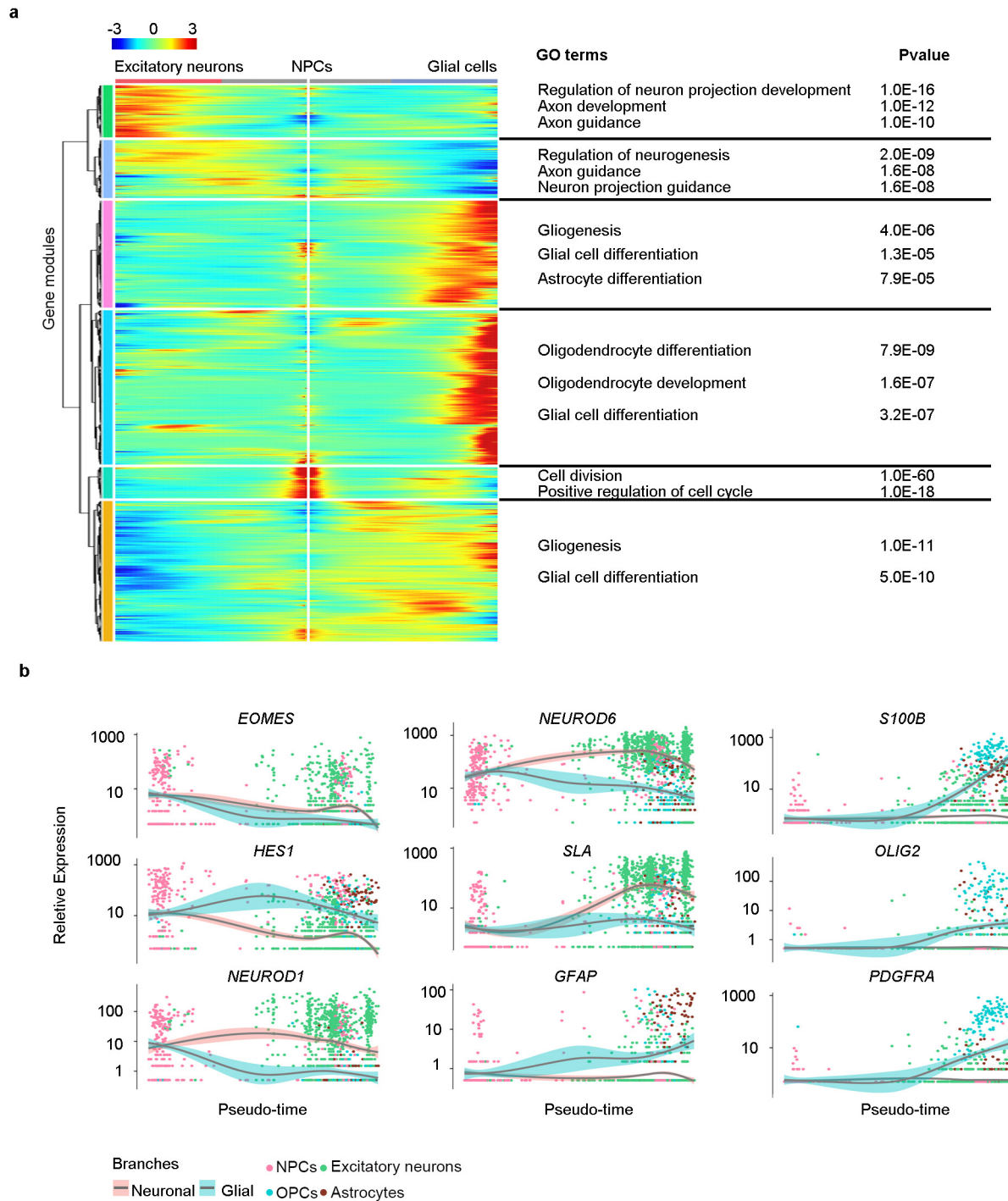
Interneurons



Astrocytes

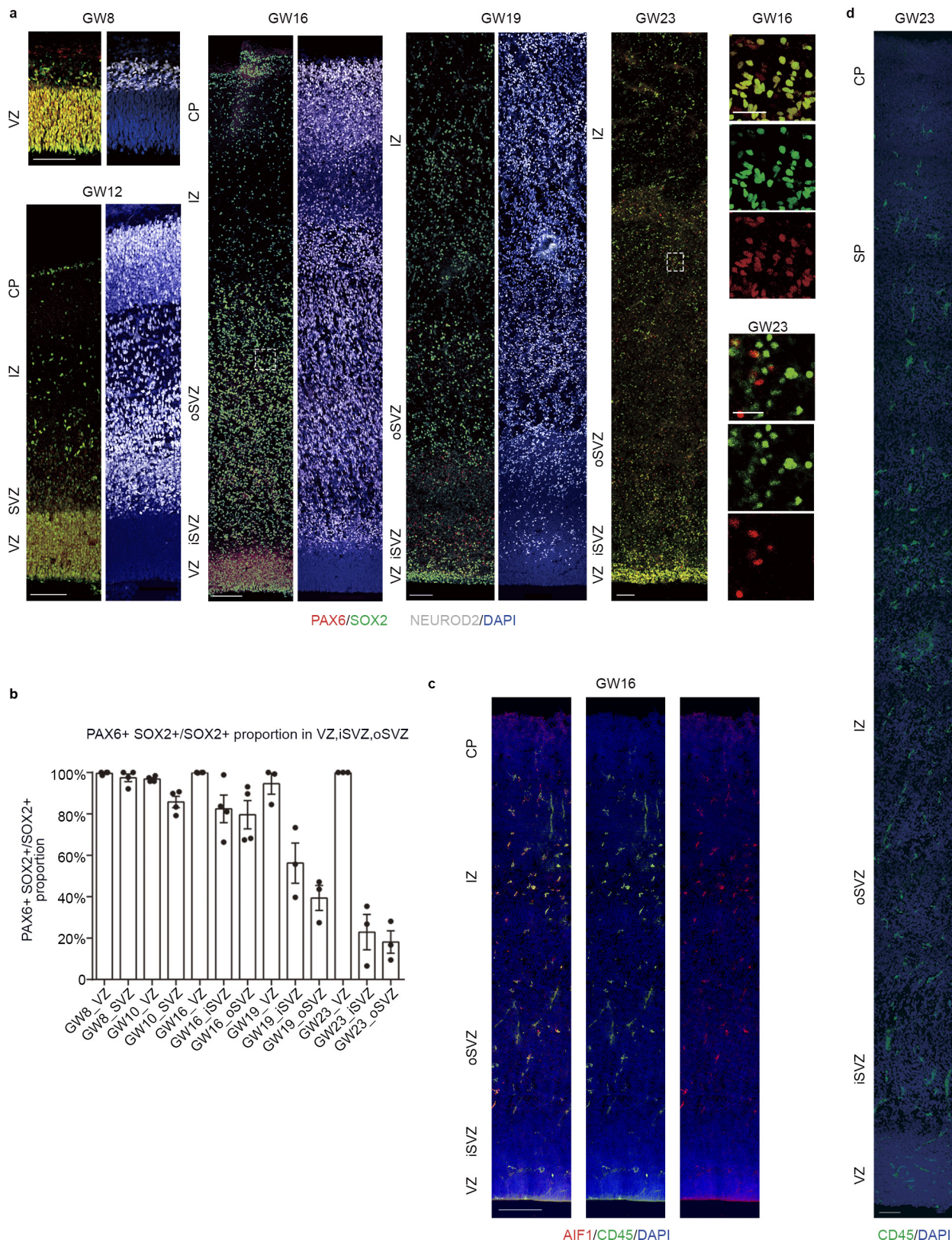


Extended Data Figure 3 | Pie chart of the distribution of the 35 subclasses of the six cell types across gestational time.



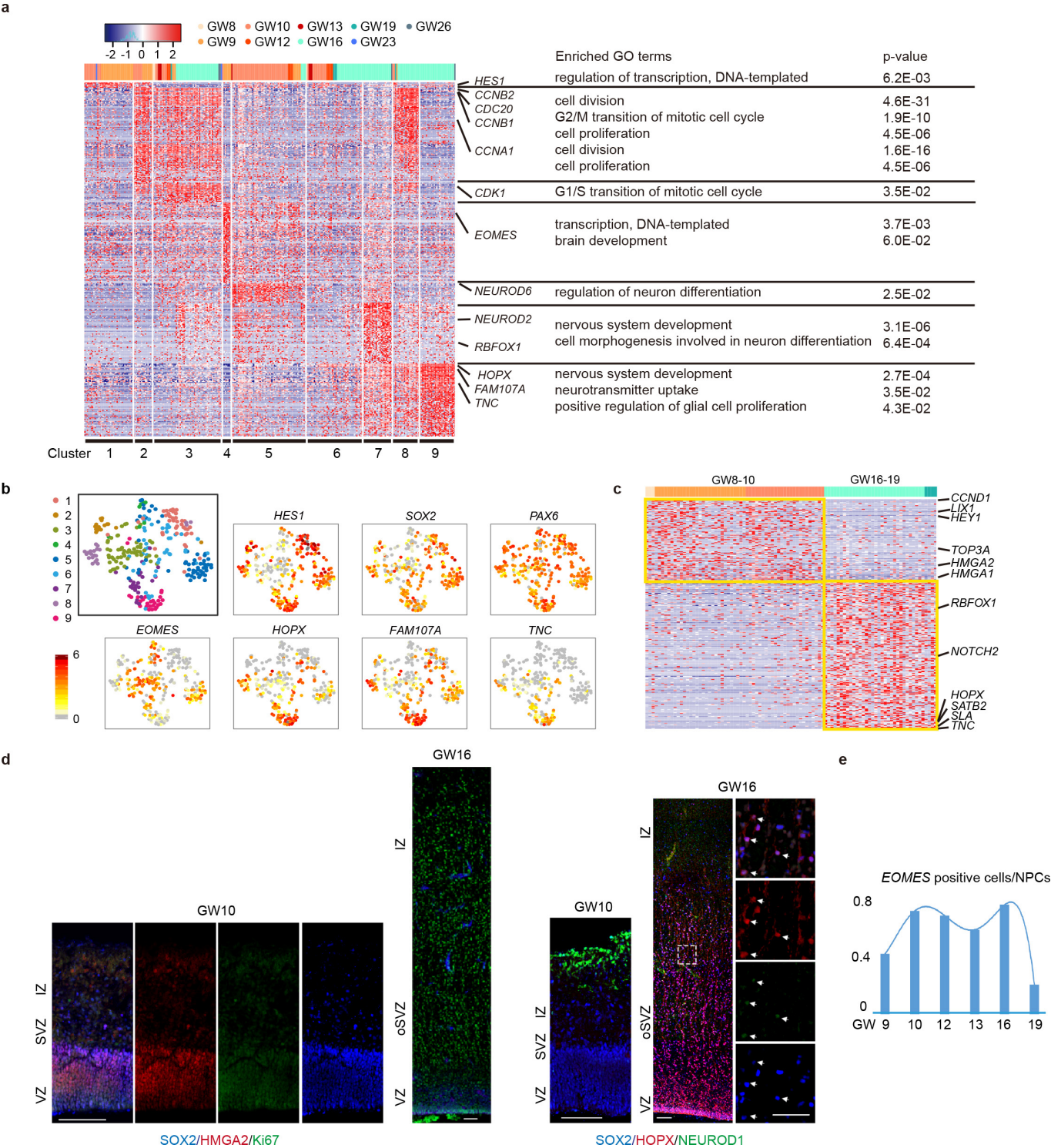
Extended Data Figure 4 | Relative expression of known markers of each cell type. a, The bifurcation of gene expression along two branches is clustered hierarchically into six modules. Gene ontology analysis of each module reflected the processes controlling neuronal and glial fate commitment. In this heat map, columns are points in pseudo-time, rows are genes and the middle (NPCs) is the beginning of pseudo-time. One

lineage goes from the middle of the heat map to the right (glial cells) while the other lineage goes to the left (excitatory neurons). $n = 1,540$ cells. **b,** The markers (*EOMES*, *HES1* for NPCs; *NEUROD1*, *NEUROD6*, *SLA* for excitatory neurons; *GFAP*, *S100B* for astrocytes; *OLIG2*, *PDGFRA* for OPCs) were ordered by Monocle analysis in pseudo-time as in Fig. 1c; the shadow indicates the confidence interval around the fitted curve.



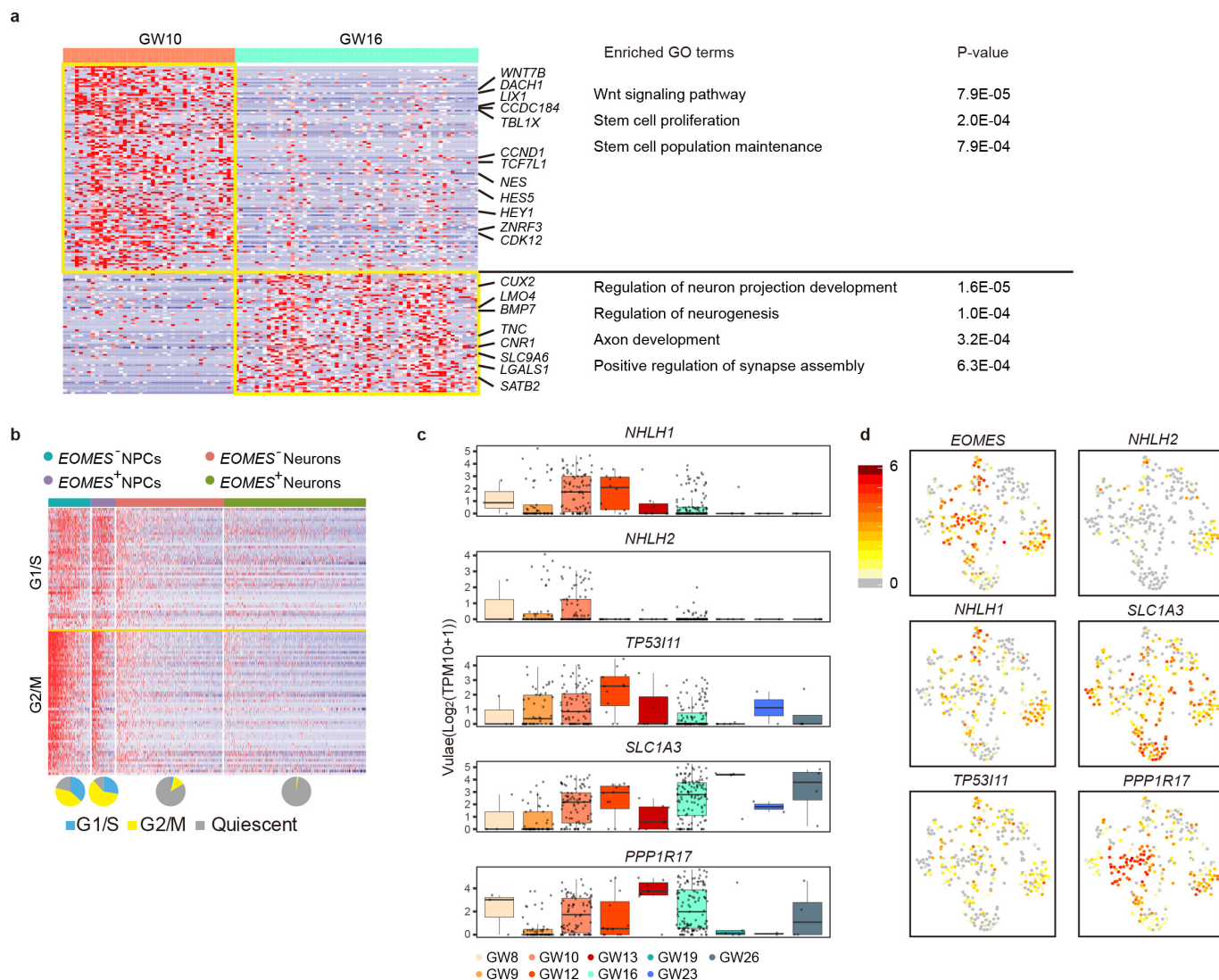
Extended Data Figure 5 | Immunostaining of neural progenitor cells and microglia in the developing PFC. **a**, Immunostaining for known markers (PAX6 and SOX2 for NPCs, NEUROD2 for excitatory neurons) at GW8, GW12, GW16, GW19 and GW23, showing the position of NPCs and excitatory neurons. Scale bars, 100 μ m (left) and 25 μ m (right). $n = 3$ for GW 19 and GW23, $n = 4$ for GW8, GW12 and GW16. **b**, Bar

charts showing the proportion of PAX6⁺SOX2⁺ cells in SOX2⁺ cells in the ventricular zone, iSVZ and oSVZ in the PFC at GW8, GW12, GW16, GW19 and GW23 relative to the images in **a**. Data are mean \pm s.e.m. **c**, Co-staining for AIF1 and CD45 to label microglia in the PFC at GW16. **d**, Immunostaining for the microglia marker CD45 in the PFC at GW23. Scale bar, 100 μ m. $n = 3$ independent replicates per gestational week.



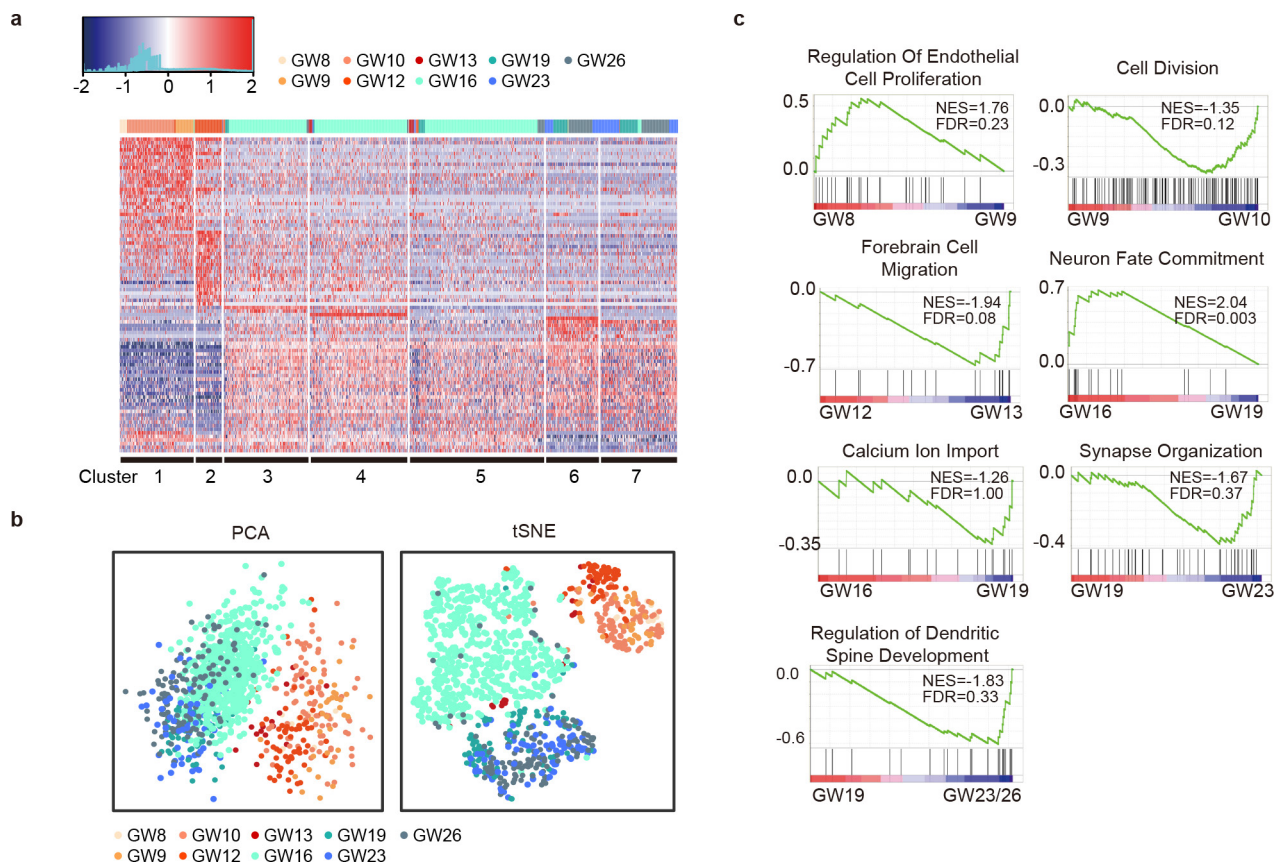
Extended Data Figure 6 | Subclasses and Monocle analysis of neural progenitor cells in the developing PFC. **a**, Heat map of differentially expressed genes of subclasses in NPCs. Expression of known genes in each type is shown on the right of each heat map panel with enriched gene ontology terms. The graph on the top shows the distribution of each subclass across gestational week. $n = 288$ cells. **b**, Visualization of nine major classes of NPCs using t -SNE (colour on the left, subtype of NPCs) with known marker expression (right and bottom: grey, no expression;

yellow-red, relative expression). **c**, Heat map of differentially expressed genes between GW8 to GW10 and GW16 to GW19 in $EOMES^+$ NPCs. The graph at the top shows the distribution of each subclass across gestational weeks. **d**, Immunostaining for different stage markers of GW10 (HMGA2) and GW16 (HOPX) in the PFC. Scale bars, 100 μ m. $n = 3$ independent replicates per gestational week. **e**, Histogram showing the ratio of $EOMES^+$ cells to all NPCs across gestational weeks quantified by RNA-seq data.



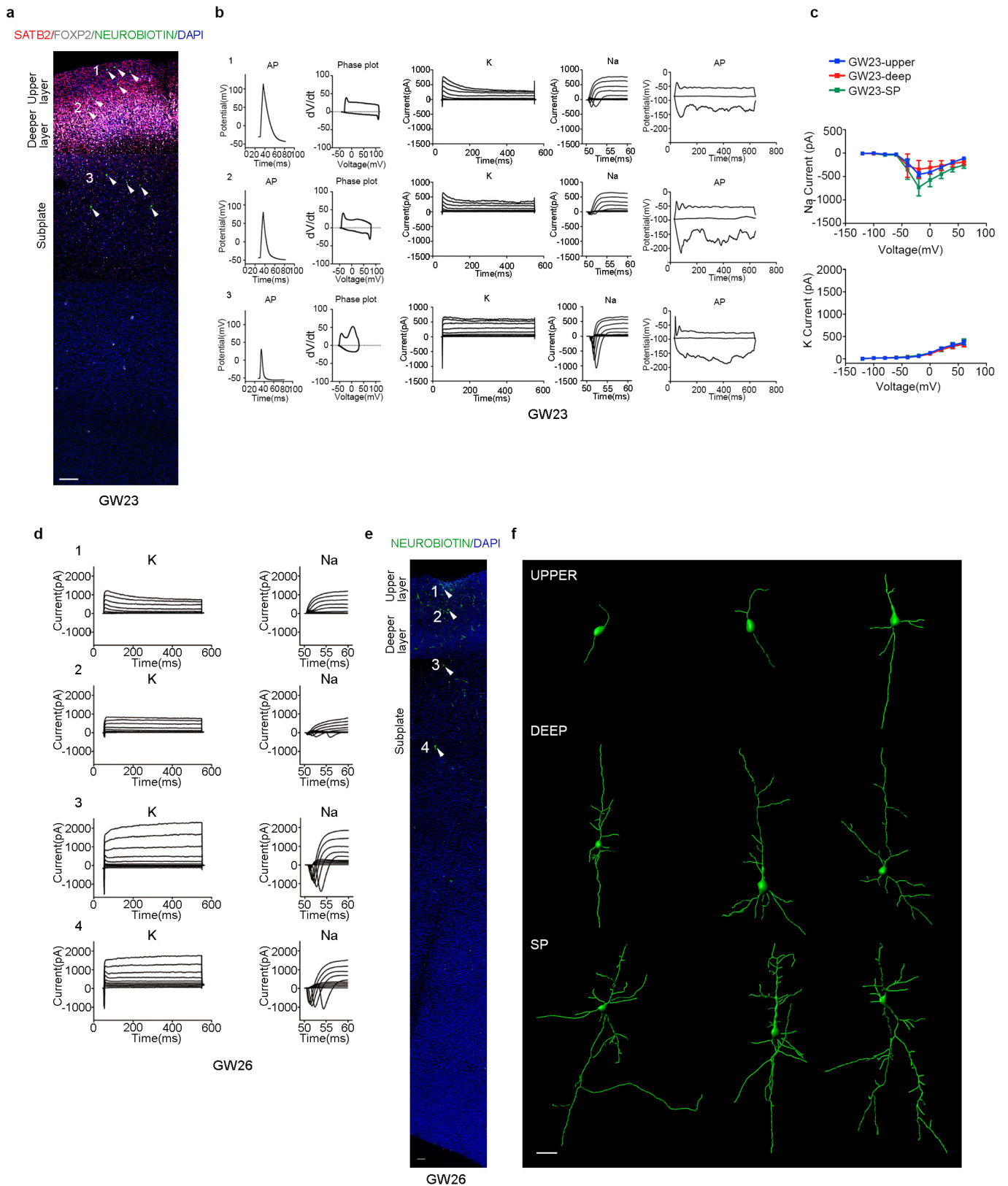
Extended Data Figure 7 | Expression of markers of IPCs in the developing PFC. a, Heat map of differentially expressed genes in *EOMES*⁺ NPCs at GW10 and GW16. Expression of known genes at each stage is shown to the right of each heat map panel with enriched gene ontology terms. *n* = 288 cells. **b,** Heat map of expression of cell-cycle genes in all NPCs and neurons. The pie chart at the bottom indicates the ratios of cell types at each cell-cycle stage. **c,** Gene expression of novel markers of IPCs

across developmental time corresponding to human cortical neurogenesis. NPCs: GW8, 3 cells; GW9, 50 cells; GW10, 84 cells; GW12, 12 cells; GW13, 7 cells; GW16, 123 cells; GW19, 5 cells; GW23, 2 cells; GW26, 4 cells. **d,** The expression of novel markers of IPCs is shown using the same layout as Extended Data Fig. 2b (grey, no expression; yellow-red, relative expression). *n* = 288 NPCs.



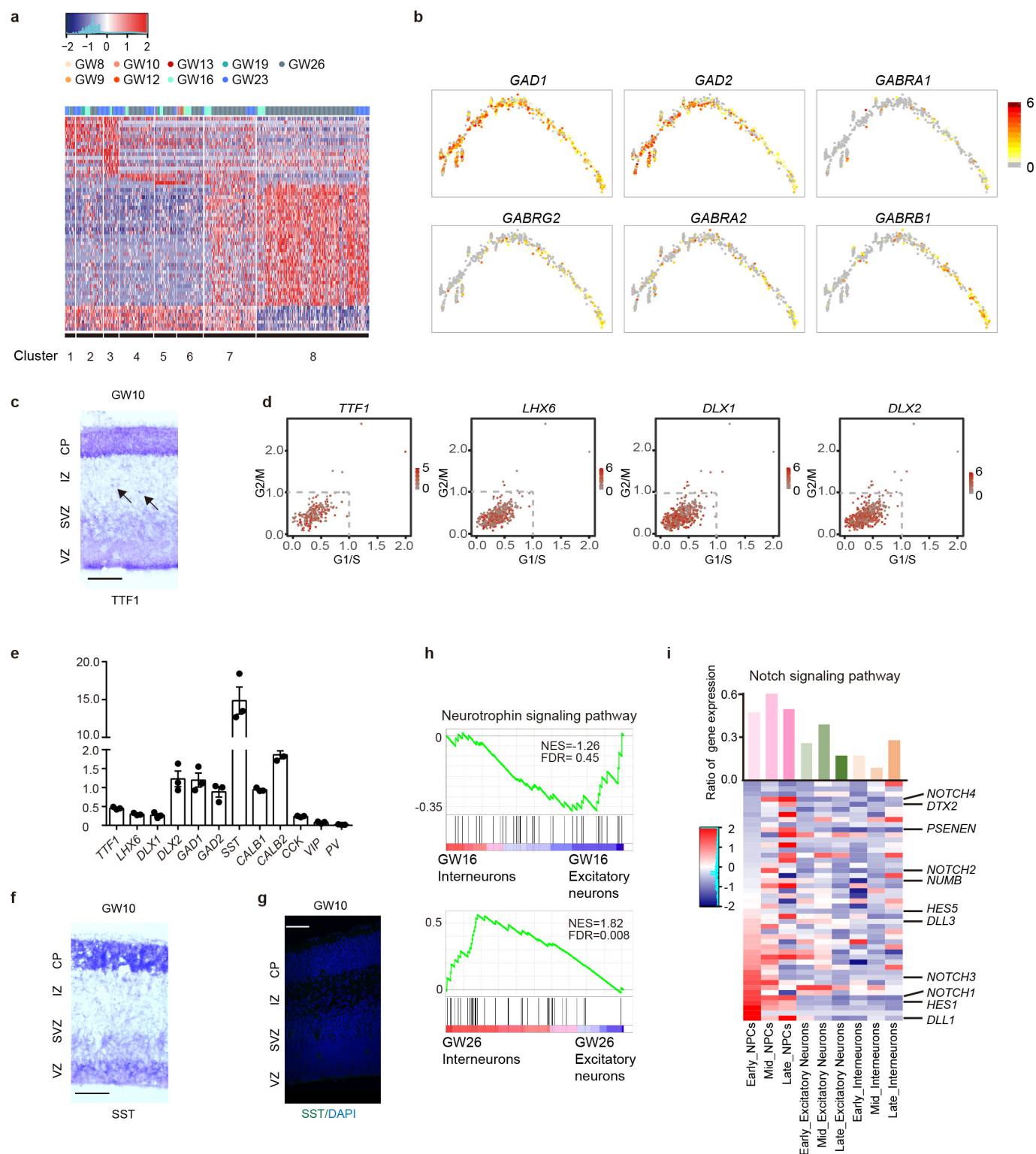
Extended Data Figure 8 | Subclasses of excitatory neurons in the developing PFC. a, Heat map of differentially expressed genes in different subtypes of excitatory neurons. The bar chart at the top shows the gestational week. **b,** Principal component analysis and *t*-SNE were used to sort excitatory neurons into subgroups. Each dot represents a single cell.

The colour shows the gestational week. **c,** GSEA shows genes related to key functions of excitatory neurons in the PFC across multiple developmental time points corresponding to human cortical neurogenesis. $n = 1,057$ excitatory neurons.



Extended Data Figure 9 | Whole-cell patch-clamp recordings in the developing PFC. a, Immunostaining after whole-cell patch-clamp recordings of the PFC at GW23. Scale bar, 100 μ m. $n = 3$ independent replicates. **b**, Whole-cell patch-clamp recordings of the PFC at GW23. **c**, Quantification of Na^+ current (top) and K^+ current (bottom) of neurons located in upper layer, deep layer or subplate of GW23 PFC. $n = 31$ neurons from three independent replicates. Data are mean \pm s.e.m.

All $P > 0.05$, two-way ANOVA. **d**, Whole-cell patch-clamp recordings of the PFC neurons at GW26. **e**, Immunostaining after whole-cell patch-clamp recordings of the PFC neurons at GW26. Scale bar, 100 μ m. **f**, Three-dimensional reconstructions of neurons are shown. Scale bar, 30 μ m. Nine representative images are shown from $n = 26$ neurons from three independent replicates.



Extended Data Figure 10 | Subclasses of interneurons and signal pathways regulating neurogenesis. **a**, Heat map of illustrated subclasses of interneurons. The bar chart on the top shows the gestational week. **b**, The expression of known markers of interneurons mapped to the Monocle analysis (Fig. 4a) shows the pseudo-time course of interneurons during development (grey, no expression; yellow–red, relative expression). **c**, *In situ* hybridization of *TTF1* shows the position of interneuron progenitor cells in PFC at GW10. Scale bar, 100 μ m. $n = 3$ independent replicates. **d**, Pattern of marker gene expression in interneuron progenitor cells mapped onto the cell cycle plot for all interneurons. Interneurons with high expression of *TTF1*, *LHX6*, *DLX1* and *DLX2* exhibit low

expression of cell cycle genes. **e**, Quantification of specific markers for interneuron progenitor cells and interneurons of PFC at GW7 by reverse transcription–PCR analysis. $n = 3$ independent replicates. **f**, *In situ* hybridization of *SST* in PFC at GW10. Scale bar, 100 μ m. **g**, Immunostaining of *SST* in PFC at GW10. Scale bar, 100 μ m. $n = 3$ independent replicates. **h**, GSEA enrichment plot of the KEGG neurotrophin signalling pathway. $n = 663$ excitatory neurons; $n = 485$ interneurons. **i**, Mean expression of Notch signalling pathway genes in NPCs, excitatory neurons and interneurons at different stages of development. The bar chart at the top represents the ratio of expressed genes to all genes in the Notch pathway.

Diffusible repression of cytokinin signalling produces endodermal symmetry and passage cells

Tonni Grube Andersen¹, Sadaf Naseer¹, Robertas Ursache¹, Brecht Wybouw^{2,3}, Wouter Smet^{2,3,4}, Bert De Rybel^{2,3,4}, Joop E.M. Vermeer^{1†} & Niko Geldner¹

In vascular plants, the root endodermis surrounds the central vasculature as a protective sheath that is analogous to the polarized epithelium in animals, and contains ring-shaped Casparian strips that restrict diffusion¹. After an initial lag phase, individual endodermal cells suberize in an apparently random fashion to produce ‘patchy’ suberization that eventually generates a zone of continuous suberin deposition². Casparian strips and suberin lamellae affect paracellular and transcellular transport, respectively. Most angiosperms maintain some isolated cells in an unsuberized state³ as so-called ‘passage cells’, which have previously been suggested to enable uptake across an otherwise-impermeable endodermal barrier^{3,4}. Here we demonstrate that these passage cells are late emanations of a meristematic patterning process that reads out the underlying non-radial symmetry of the vasculature. This process is mediated by the non-cell-autonomous repression of cytokinin signalling in the root meristem, and leads

to distinct phloem- and xylem-pole-associated endodermal cells. The latter cells can resist abscisic acid-dependent suberization to produce passage cells. Our data further demonstrate that, during meristematic patterning, xylem-pole-associated endodermal cells can dynamically alter passage-cell numbers in response to nutrient status, and that passage cells express transporters and locally affect the expression of transporters in adjacent cortical cells.

For more than a century, it has been known that angiosperm roots display interspersed passage cells in their suberized endodermis⁴. In monocots, these cells remain thin-walled and unsuberized for many months⁴, which suggests that passage cells represent a stable cell fate. In *Arabidopsis*, passage cells have only sporadically been mentioned and the scarce experiments that have addressed the function of these cells have mostly been correlative^{3,5}. Although the molecular basis of passage cell development is as yet unknown, suberization in *Arabidopsis* follows a stereotypic pattern² that has recently been shown to be highly

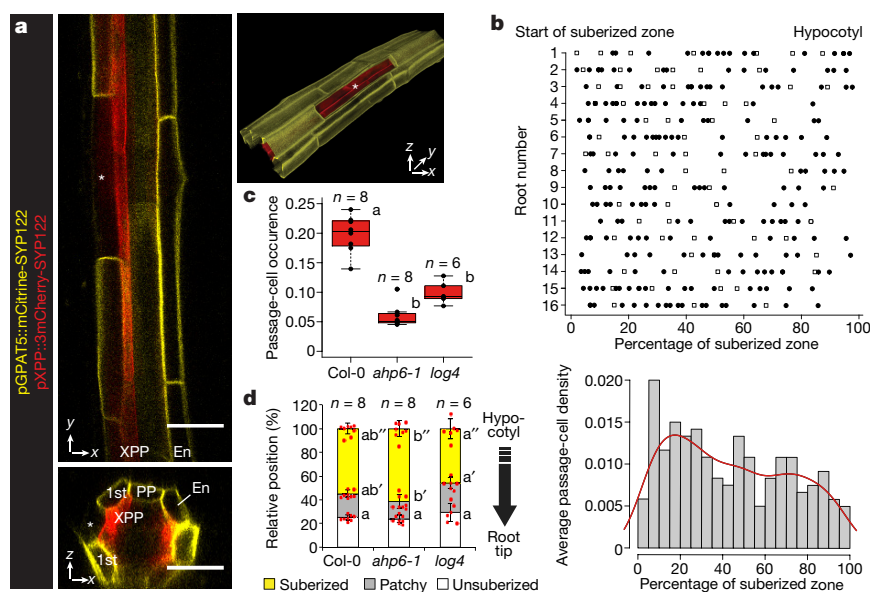


Figure 1 | Presence and distribution of passage cells in *Arabidopsis*.

a, Representative suberized endodermis and xylem-pole pericycle visualized by a suberin (*GPAT5*) and xylem-pole pericycle (XPP) reporter, using plasma-membrane-localized mCitrine-SYP122 or 3mCherry-SYP122 reporters, respectively. Asterisk, passage cell. **b**, Top, scoring of lateral root primordia (white squares) and passage cells (black circles) from the start of the fully suberized zone (0%) to the hypocotyl junction (100%). Bottom, 5% binning of data from top panel (grey bars), and trend of passage-cell density along the suberized zone (red line) **c**, Passage-cell occurrence in Col-0 plants, and *ahp6-1* and *log4* mutants. **d**, Suberization in Col-0 plants, and *ahp6-1* and *log4* mutants. En, endodermis;

PP, phloem pole. Bar graphs represent mean \pm s.d. and boxplot centres show median. In all graphs, dots represent individual data points. For all stacked graphs, there are three measurements per root: unsuberized zone, white; patchy zone, grey; and suberized zone, yellow. Individual letters show significantly different groups according to a post hoc Bonferroni-adjusted paired two-sided *t*-test. The prime symbols indicate that *t*-tests were performed on measurements within identical zones, that is two primes for suberized, one prime for patchy and none for unsuberized. For more information on data plots, see Methods. The image in **a** is representative of five independent lines. *n*, independent biological samples. For individual *P* values, see Supplementary Table 2. Scale bars, 25 μ m.

¹Department of Plant Molecular Biology, University of Lausanne, 1015 Lausanne, Switzerland. ²Ghent University, Department of Plant Biotechnology and Bioinformatics, 9052 Ghent, Belgium.

³VIB Center for Plant Systems Biology, 9052 Ghent, Belgium. ⁴Wageningen University, Laboratory of Biochemistry, 6708 WE Wageningen, The Netherlands.

[†]Present address: Department of Plant and Microbial Biology, University of Zürich, 8008 Zürich, Switzerland.

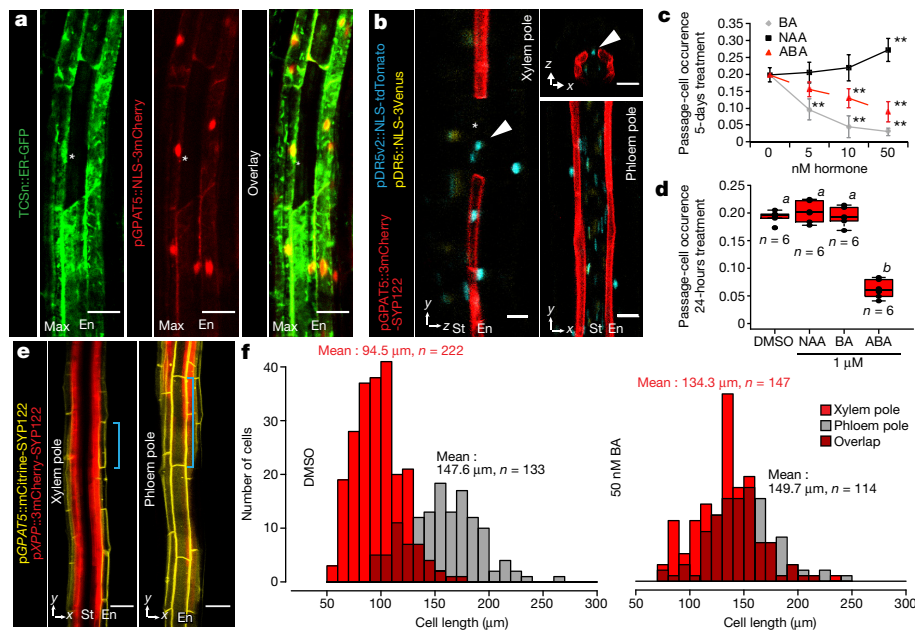


Figure 2 | Cytokinin and auxin regulate endodermal patterning and passage cell formation. **a**, Representative image depicting expression of the cytokinin response marker *TCSn* (ER–GFP, green; ER, endoplasmic reticulum) or the suberin reporter *GPAT5* (NLS–3mCherry, red; NLS, nuclear localization signal) in fully suberized endodermis. **b**, Expression of auxin signalling reporter *DR5v2* (NLS–tdTomato, blue), or *DR5* (NLS–3mCherry–SYP122, red) in fully suberized endodermis. Red dots, individual data points. **c**, **d**, Occurrence of passage cells in seedlings germinated on indicated hormones (**c**) or after 24-h hormone incubation (**d**). Dimethyl sulfoxide (DMSO), mock treatment. Black dots, individual data points. **e**, Optical sections through suberized endodermis. Suberin and xylem-pole pericycle highlighted as in Fig. 1a. Blue lines, length of a single cell in xylem or

phloem pole. **f**, Distribution of suberized endodermal cell length in xylem or phloem pole of plants germinated on mock (DMSO) or cytokinin plates. Numbers depict average lengths of xylem-pole (red) or phloem-pole (grey) endodermal cells. Asterisks indicate passage cells and arrowheads indicate passage-cell nuclei. BA, benzyl adenine; NAA, naphthalene acetic acid; Max, maximum projection; St, stele. Boxplot centres show median. Statistically significant differences between groups were tested using a post hoc Bonferroni-adjusted paired two-sided *t*-test. In **a**, **b** and **e**, image is representative of eight independent lines. *n*, independent biological samples. In **c**, ***P* < 0.01, two-tailed *t*-test. Dots represent mean, error bars are s.d. and *n* = 25 independent biological replicates for each treatment. For more information regarding data plots, see Methods. For individual *P* values, see Supplementary Table 2. Scale bars, 25 μ m.

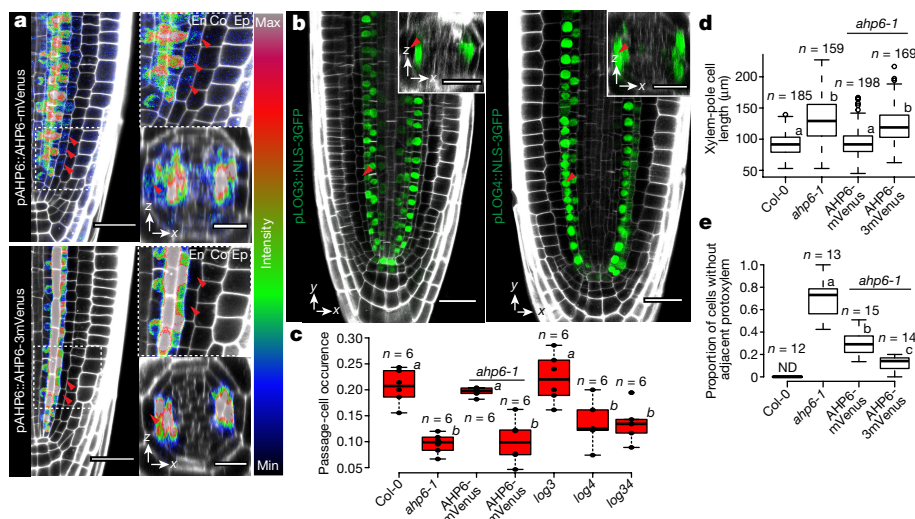


Figure 3 | Spatially restricted cytokinin repression and production underlie passage cell formation. **a**, Expression of pAHP6::AHP6-mVenus or pAHP6::AHP6-3mVenus from a previously published line¹⁹ in the root meristem. Figures represent longitudinal and transverse optical sections through the xylem pole of the root meristematic zone. **b**, *pLOG3* and *pLOG4* reporter activity (NLS–3GFP, green) in the root meristem. The lines that we used have previously been published⁸. **c**, Occurrence of passage cells in *ahp6-1* and lines complemented with AHP6 fused to a single or triple mVenus protein, driven from *AHP6* promoter. Dots, individual data points. **d**, Length of xylem-pole endodermal cells in the suberized zone of *ahp6-1* and in lines complemented with AHP6 fusions.

Dots, outliers. **e**, Quantification of xylem defects in *ahp6* mutant lines (for details, see Methods). Co, cortex; Ep, epidermis; ND, not detected. Red arrowheads, xylem-pole endodermal cells. Boxplot centres show the median. Statistically significant differences between groups were tested using a post hoc Bonferroni-adjusted paired two-sided *t*-test. For more information on data plots, see Methods. In **a** and **b**, images are representative of previously published lines^{8,19}. *n*, independent biological samples. In **d**, *n* represents individual measurements across 16 independent biological samples. For individual *P* values, see Supplementary Table 2. Scale bars, 25 μ m.

responsive to many stress conditions, mediated by abscisic acid (ABA) and ethylene². In the zone of continuous suberization, we found individual cells that lacked suberin deposition (Fig. 1a), which was reliably paralleled by a live marker for suberization² (Extended Data Fig. 1a–c). Using a marker for the xylem-pole pericycle (Extended Data Fig. 1d), we demonstrate a close association between these suberin-lacking cells and the xylem pole (Extended Data Fig. 1f); this close association is a second defining feature of passage cells³. Similar to other angiosperms, suberization in *Arabidopsis* begins above the phloem pole, approximately four cells closer to the root tip than above the xylem pole³ (Extended Data Fig. 1g, h). Although passage cells appear randomly along the longitudinal axis and are not correlated with sites of lateral root emergence, they are sometimes clustered and have a tendency to decrease towards the hypocotyl (Fig. 1b and Extended Data Fig. 1e). To understand the mechanism that determines the association between passage cells and the xylem pole, we investigated mutants of genes involved in xylem patterning. Two cytokinin-related mutants, *ahp6-1* (also known as *hp6-1*) and *log4*, showed reduced numbers of passage cells without affecting overall suberization (Fig. 1c, d). AHP6 attenuates cytokinin responses⁶, and LOG4 is involved in cytokinin biosynthesis^{7,8}. Auxin–cytokinin interactions are essential to establish the bisymmetric pattern of phloem and xylem poles^{9,10}. The preferential accumulation of auxin in xylem precursors is thought to lead to the expression of AHP6 and LOG4, turning these cells into a cytokinin-refractory cytokinin source. Higher levels of cytokinin signalling in neighbouring cells then induces procambium and/or phloem pole cells⁸, which in turn usher more auxin towards xylem precursors and thereby establish complementary domains of auxin and cytokinin perception¹⁰. We hypothesized that these bisymmetric signalling domains also cause the association of passage cells with the xylem pole.

Using a cytokinin-response marker¹¹, we observed responses in the suberized root zone. Although cytokinin response markers were observed most frequently in the pericycle, they were also observed in the suberized endodermis (Fig. 2a and Extended Data Fig. 2b) but not in passage cells, which indicates an absent or attenuated cytokinin response in the latter cells (Fig. 2a). By observing the transcriptional expression patterns of most genes of the *Arabidopsis* response regulator (ARR) family, which contains negative (A-type) and positive (B-type) transcriptional regulators of cytokinin signalling^{12–14}, we found that repressive A-type *ARR3* and *ARR6*, as well as the B-type *ARR14*, were expressed in passage cells; suberized endodermal cells, however, showed no expression of A-type ARRs (Extended Data Fig. 2c, d). This demonstrates that passage cells have a distinct set of cytokinin-response regulators and possibly explains their attenuated cytokinin response. Our inability to detect ARRs in suberized endodermis might be due to their low abundance in these cells or the fact that not all ARRs were represented in our marker set. Using the standard DR5 reporter, we detected auxin signalling only in vasculature and tissues surrounding lateral root primordia (Fig. 2b and Extended Data Fig. 2a). Using an improved version of the auxin reporter¹⁵, however, we observed additional signals that were restricted to xylem-pole endodermal cells but were not exclusive to passage cells (Fig. 2b). Passage cells are therefore associated with differential auxin and cytokinin responses within the circumference of the late endodermis.

Germinating seedlings on auxin increased the number of passage cells, but only at concentrations that also affected root growth (Fig. 2c and Extended Data Fig. 3). By contrast, cytokinin decreased passage-cell numbers even at concentrations that did not affect root growth (Fig. 2c and Extended Data Fig. 3). ABA strongly promotes endodermal suberization, and caused enhanced and precocious deposition of suberin². As both ABA and cytokinin decreased the number of passage cells (Fig. 2c), we investigated how the two hormones might be connected. Seedling transfer to ABA-containing, but not to cytokinin-containing, plates for 24 h led to passage-cell closure (Fig. 2d), which was already observable after 9 h (Extended Data Fig. 4). This indicates that ABA can act during the late stages of endodermal

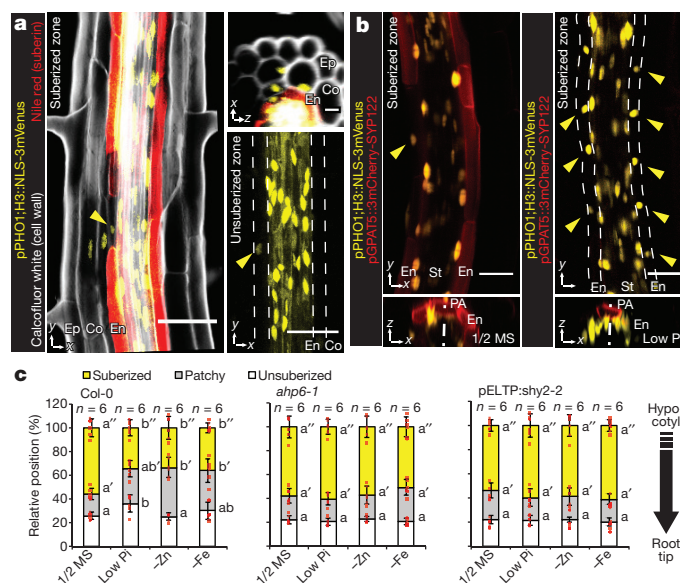


Figure 4 | Passage cell-associated expression of *PHO1;H3*. **a**, *PHO1;H3* reporter activity (NLS-3mVenus) in the suberized and unsuberized zones of differentiated endodermis. Cell walls (grey) visualized by calcofluor white and suberin (red) using Nile red. **b**, *PHO1;H3* expression in suberized zone of seedlings germinated on standard medium (1/2 Murashige and Skoog medium, 1/2 MS) or 10 μ M inorganic phosphate (low Pi). Suberized endodermis cells highlighted by *GAT5* expression. Yellow arrowheads, passage-cell nuclei. **c**, Endodermal suberization in Col-0, *ahp6-1* and lines with repressed auxin signalling (through expression of *shy2-2*) in differentiated endodermis grown on standard medium, low Pi, or zinc (–Zn) or iron-deficient (–Fe) medium. Red dots, individual data points. For all stacked graphs, there are three measurements per root: unsuberized zone, white; patchy zone, grey; and suberized zone, yellow. PA, phloem axis. Bar graphs represent mean \pm s.d. Statistically significant differences between groups were tested using a post hoc Bonferroni-adjusted paired two-sided *t*-test. For more information regarding data plots, see Methods. In **a** and **b**, images are representative of 12 independent lines. *n*, independent biological samples. For individual *P* values, see Supplementary Table 2. Scale bars, 25 μ m.

development, whereas cytokinin affects meristematic patterning events. We observed that xylem-pole endodermal cell length is about half that of phloem-pole cells (Fig. 2e, f). Similar dimorphisms have previously been described for other species⁴ and could arise from a cytokinin-dependent delay of exit from the division zone in xylem-pole-associated cells^{16,17}.

Increasing concentrations of cytokinin increased the average length of xylem-pole-associated endodermal cells, which then approached the length of phloem-pole cells (Fig. 2f and Extended Data Fig. 5f). Consistent with its antagonistic action, auxin decreased the length of xylem-pole cells (Extended Data Fig. 5e); ABA did not affect endodermal cell length (Extended Data Fig. 5d). These data indicate that cytokinin causes a difference between xylem- and phloem-pole-associated endodermis in the transition zone; we use this difference as an early read-out of bisymmetric patterning within the endodermis (Extended Data Fig. 5g).

To test whether cytokinin and auxin act directly in the endodermis, we specifically overexpressed cytokinin- or auxin-signalling suppressors in all differentiating endodermal cells. Cytokinin inhibition caused an almost complete absence of suberization, as if all endodermal cells had acquired a passage-cell identity (Extended Data Fig. 5a–c and Extended Data Fig. 6a–d). This suppression of suberization could not be antagonized by ABA, supporting a model in which cytokinin signalling determines the responsiveness of endodermal cells to ABA (Extended Data Fig. 6e). Suberization persisted around lateral root emergence sites, which suggests that in these areas suberization is independent of cytokinin (Extended Data Fig. 5a). When inhibiting endodermal auxin signalling, we observed decreased passage-cell numbers (Extended Data Fig. 5b, c). We added a temporal control to these manipulations by using an oestradiol-inducible expression system¹⁸. A 29-h oestradiol induction

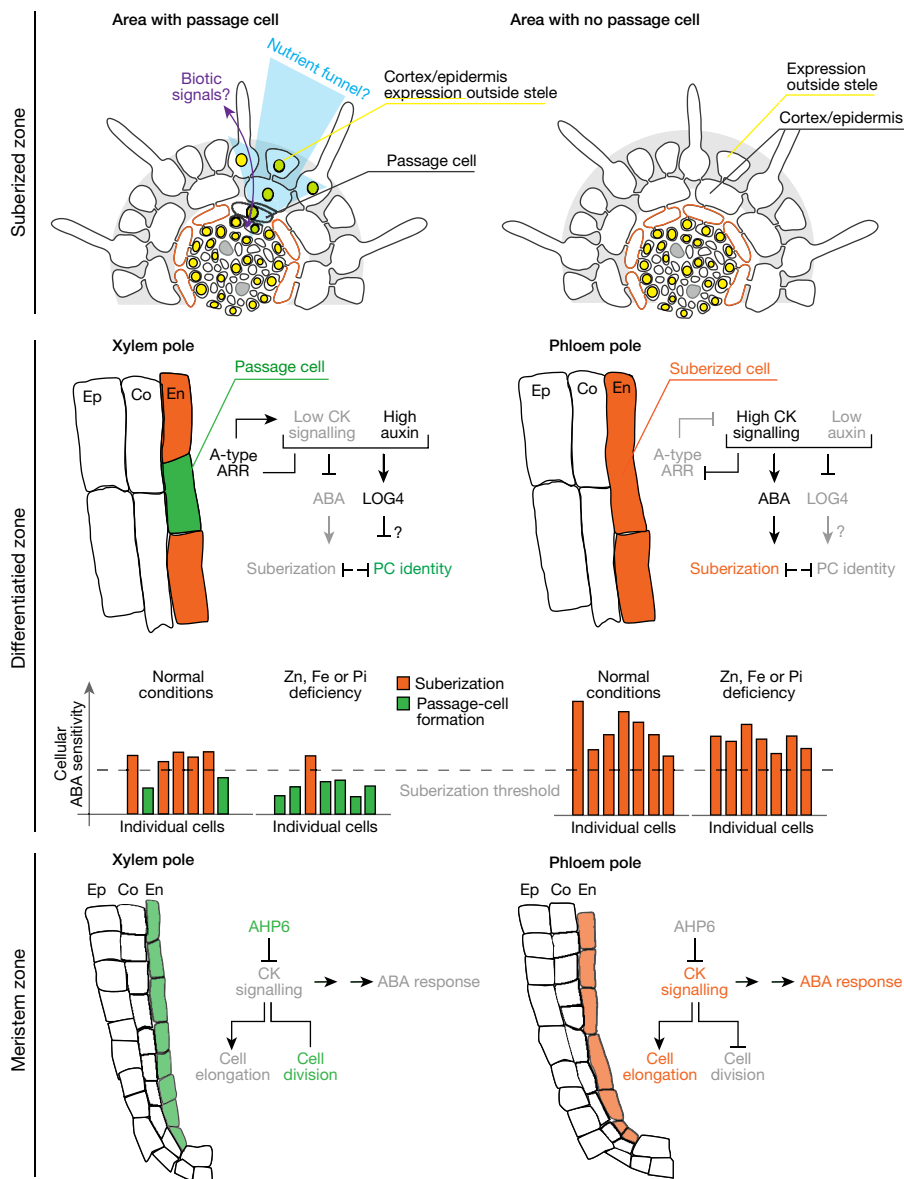


Figure 5 | Overview of endodermis circumferential patterning and passage cell formation. In the meristem zone, endodermal cells above the incipient xylem pole are exposed to AHP6 produced in the stele, which inhibits cytokinin signalling. This leads to prolonged division and thus shorter endodermal cells than in the phloem pole. Another output should be an overall lower response to ABA, through an as-yet unknown mechanism. In the differentiation zone, we hypothesize that the lower ABA response in xylem-pole endodermis stochastically causes some cells to be below the threshold for suberization. The unsuberized cells maintain low cytokinin signalling by expression of A-type ARR elements and also serve as—possibly auxin-dependent—cytokinin producers through the expression of LOG4. The status of ABA signalling is modulated by nutrient status, and thus passage cells can be closed given sufficiently high levels of ABA. In the fully suberized zone, the unsuberized xylem-pole-associated endodermal cells (passage cells) express specific genes with a domain that can spread to the cortex and/or epidermis, and serve to increase the area for uptake and/or to exchange signals with the environment (such as signals for biotic interactions). Nutrient stresses (zinc, iron or phosphate deficiencies) would increase the resistance of xylem-pole-associated endodermis cells towards ABA and lead to an increase in unsuberized cells, enabling increased transport-related gene expression. CK, cytokinin; PC, passage cell.

of AHP6–GFP did not affect suberization that was already established, or established passage cells, thereby confirming cytokinin application experiments (Extended Data Fig. 7). By contrast, auxin-repressor induction for 29 h reduced passage-cell numbers (Extended Data Fig. 7), which suggests that auxin signalling is also required to maintain passage-cell fate. Repressing ABA for 29 h led to an almost complete disappearance of suberin (Extended Data Fig. 7), which suggests that the strong suppression of ABA signalling interferes with maintenance of suberization, possibly by de-repressing ethylene signalling². Having established the direct endodermal action of cytokinin, auxin and ABA in passage-cell formation, we sought to understand how spatial differences in cytokinin presence or perception might arise.

One regulator of cytokinin perception in the xylem pole is AHP6, the presence of which interferes with phospho-transfer reactions from receptors towards transcriptional regulators⁶. Although its transcription is confined to the stele, we found that a complementing AHP6–mVenus fusion diffuses into endodermal cells above the xylem pole (Fig. 3a and Extended Data Fig. 8a) where it may attenuate cytokinin signalling. To establish the relevance of this observation, we used a functional, non-mobile triple-mVenus AHP6 fusion¹⁹. Only the mobile single-mVenus fusion rescued passage-cell number and xylem-pole length of endodermal cells in *ahp6-1* mutants (Fig. 3c, d and Extended Data Fig. 8b). This was not due to the lower activity of the triple-m

Venus fusion, as the xylem patterning defects of *ahp6-1* were rescued to an even higher extent when a triple- rather than single-mVenus fusion was used (Fig. 3e). Thus, circumferential endodermal patterning and passage-cell differentiation relies on the movement of AHP6 from the stele into the endodermis.

In the stele, the cytokinin biosynthetic enzymes LOG3 and LOG4 turn xylem precursors into a source of cytokinin, which enhances signalling in neighbouring cells⁸. Interestingly, *log4*, but not *log3*, mutants showed lower passage-cell numbers (Fig. 3c and Extended Data Fig. 8g). *log4 ahp6-1* double mutants did not display a further reduction in passage cells (Extended Data Fig. 8g), which suggests that AHP6 and LOG4 act in a single pathway. The specificity of LOG4 can be explained by LOG3 transcription being restricted to the stele, whereas LOG4 is mainly expressed in xylem-pole-associated endodermal cells⁸ (Fig. 3b). LOG4–GFP expression in differentiating, but not meristematic, endodermal cells rescued passage-cell numbers (Extended Data Fig. 8f), suggesting that LOG4 maintains passage-cell differentiation rather than being required for specification. Additionally, in *log4* mutants the length of xylem-pole endodermis is not affected (Extended Data Fig. 8b). Both AHP6 and LOG4 expression are reduced by cytokinin application⁶ (Extended Data Fig. 8a, c) and this effect could explain the high sensitivity of passage-cell differentiation towards cytokinin. Combining *ahp6-1* with late endodermis-specific inhibition of auxin signalling further reduced

passage-cell differentiation (Extended Data Fig. 8e, h), which suggests that—in the absence of cytokinin repression—local auxin perception can partially maintain passage cells. None of the investigated mutants showed severe root developmental defects, although *ahp6-1* showed a slight reduction in lateral root emergence (Extended Data Fig. 3b).

The absence of suberization in passage cells could generate privileged sites for transport and communication. Despite the relatively small surfaces of passage cells, the uptake of some nutrients has previously been found to correlate with passage-cell numbers^{20–22} and passage cells might include transporters that would be absent in suberized cells. Indeed, the phosphate efflux protein PHO1 was reported to be expressed in both stele and xylem-pole-associated endodermal cells²². To expand on this finding, we generated sensitive, triple-mVenus-based transcriptional reporter lines for the entire PHO1 family. Besides their expression in the stele, we found PHO1 and some homologues to be specifically expressed in passage cells (Fig. 4a and Extended Data Figs 9a–c). We additionally observed clusters of cortical and epidermal expression for many of these transport-mediating genes (Fig. 4a and Extended Data Fig. 9c). Counting cluster occurrence revealed their clear spatial association with passage-cell presence in the endodermis (Extended Data Fig. 9a). The association of cortical and/or epidermal expression of PHO1 family members with underlying passage cells could arise from stele-derived signals that exit through passage cells (Fig. 4a), and possibly funnel nutrients or biotic signals from epidermis towards xylem (Fig. 5). A similar role in communication has previously been proposed for hypodermal passage cells²³.

Expression of PHO1 family members provides a positive definition of passage cells. PHO1;H3, the most easily visualized member of the family, shows expression in individual endodermal cells before the onset of suberization, which suggests that the specification of passage cells precedes suberization (Fig. 4a). We therefore used PHO1;H3 to assess whether the suppression of suberization by cytokinin correlates with expansion of the expression pattern of this gene. We found that endodermal cytokinin suppression leads to the general expression of PHO1;H3 in the endodermis, supporting the proposition that endodermal cells acquire passage-cell features on cytokinin repression (Extended Data Fig. 10b). Suppression of ABA signalling only expanded PHO1;H3 expression within xylem-pole endodermis, despite an equally strong suppression of suberization in all endodermal cells (Extended Data Figs 7, 10c), which again supports the notion that ABA acts at a later stage of endodermal development.

Finally, we investigated changes in PHO1;H3 expression under physiological stress conditions. Consistent with its putative role in phosphate transport, we found that phosphate deficiency suppressed suberization specifically in xylem-pole endodermis, and expanded PHO1;H3 expression into these cells (Fig. 4b). This response was abrogated in lines with enhanced cytokinin or suppressed auxin signalling (Fig. 4c). Expansion of PHO1;H3 expression in xylem-pole endodermis was similarly observed under conditions of zinc and iron deficiency, which have previously been shown to decrease suberization² and to enhance PHO1;H3 expression²⁴ (Fig. 4c and Extended Data Fig. 10d). This suggests that PHO1;H3 expansion is the result of an expansion of passage-cell occurrence, rather than a specific response to phosphate deficiency. qPCR analysis of PHO1;H3 expression corroborated our PHO1;H3 promoter fusion results (Extended Data Fig. 10e).

Our findings—that two endodermal cell types co-exist within roots, and possess distinct responsiveness to nutrients and hormones as well as different uptake and sensing potentials—have notable implications for current models of nutrient uptake in plants. Furthermore, the influence of isolated passage cells on neighbouring cells might explain how the small surfaces of these evolutionarily conserved cells could have important roles in nutrient transport or sensing (Fig. 5).

Online Content Methods, along with any additional Extended Data display items and Source Data, are available in the online version of the paper; references unique to these sections appear only in the online paper.

Received 1 August 2017; accepted 7 February 2018.

Published online 14 March 2018.

- Geldner, N. The endodermis. *Annu. Rev. Plant Biol.* **64**, 531–558 (2013).
- Barberon, M. *et al.* Adaptation of root function by nutrient-induced plasticity of endodermal differentiation. *Cell* **164**, 447–459 (2016).
- Peterson, C. A. & Enstone, D. E. Functions of passage cells in the endodermis and exodermis of roots. *Physiol. Plant.* **97**, 592–598 (1996).
- Kroemer, K. *Wurzelhaut, Hypodermis Und Endodermis Der Angiospermenwurzel (Bibl. Bot. 59)* (Erwin Nägele, 1903).
- Wu, H., Jaeger, M., Wang, M., Li, B. & Zhang, B. G. Three-dimensional distribution of vessels, passage cells and lateral roots along the root axis of winter wheat (*Triticum aestivum*). *Ann. Bot.* **107**, 843–853 (2011).
- Mähönen, A. P. *et al.* Cytokinin signaling and its inhibitor AHP6 regulate cell fate during vascular development. *Science* **311**, 94–98 (2006).
- Kurakawa, T. *et al.* Direct control of shoot meristem activity by a cytokinin-activating enzyme. *Nature* **445**, 652–655 (2007).
- De Rybel, B. *et al.* Integration of growth and patterning during vascular tissue formation in *Arabidopsis*. *Science* **345**, 1255215 (2014).
- De Rybel, B., Mähönen, A. P., Helariutta, Y. & Weijers, D. Plant vascular development: from early specification to differentiation. *Nat. Rev. Mol. Cell Biol.* **17**, 30–40 (2016).
- Bishopp, A. *et al.* A mutually inhibitory interaction between auxin and cytokinin specifies vascular pattern in roots. *Curr. Biol.* **21**, 917–926 (2011).
- Zürcher, E. *et al.* A robust and sensitive synthetic sensor to monitor the transcriptional output of the cytokinin signaling network in planta. *Plant Physiol.* **161**, 1066–1075 (2013).
- Hwang, I. & Sheen, J. Two-component circuitry in *Arabidopsis* cytokinin signal transduction. *Nature* **413**, 383–389 (2001).
- Sakai, H. *et al.* ARR1, a transcription factor for genes immediately responsive to cytokinins. *Science* **294**, 1519–1521 (2001).
- Mason, M. G. *et al.* Multiple type-B response regulators mediate cytokinin signal transduction in *Arabidopsis*. *Plant Cell* **17**, 3007–3018 (2005).
- Liao, C.-Y. *et al.* Reporters for sensitive and quantitative measurement of auxin response. *Nat. Methods* **12**, 207–210 (2015).
- Lavrekha, V. V., Pasternak, T., Ivanov, V. B., Palme, K. & Mironova, V. V. 3D analysis of mitosis distribution highlights the longitudinal zonation and diarch symmetry in proliferation activity of the *Arabidopsis thaliana* root meristem. *Plant J.* **92**, 834–845 (2017).
- Ioio, R. D. *et al.* A genetic framework for the control of cell division and differentiation in the root meristem. *Science* **322**, 1380–1384 (2008).
- Siligato, R. *et al.* Multisite gateway-compatible cell type-specific gene-inducible system for plants. *Plant Physiol.* **170**, 627–641 (2016).
- Besnard, F. *et al.* Cytokinin signalling inhibitory fields provide robustness to phyllotaxis. *Nature* **505**, 417–421 (2014).
- Harrison-Murray, R. S. & Clarkson, D. T. Relationships between structural development and the absorption of ions by the root system of *Cucurbita pepo*. *Planta* **114**, 1–16 (1973).
- Clarkson, D. T., Sanderson, J. & Russell, R. S. Ion uptake and root age. *Nature* **220**, 805–806 (1968).
- Hamburger, D., Rezzonico, E., MacDonald-Comber Petétot, J., Somerville, C. & Poirier, Y. Identification and characterization of the *Arabidopsis* PHO1 gene involved in phosphate loading to the xylem. *Plant Cell* **14**, 889–902 (2002).
- Sasse, J. *et al.* Asymmetric localizations of the ABC transporter PaPDR1 trace paths of directional strigolactone transport. *Curr. Biol.* **25**, 647–655 (2015).
- Khan, G. A. *et al.* Coordination between zinc and phosphate homeostasis involves the transcription factor PHR1, the phosphate exporter PHO1, and its homologue PHO1;H3 in *Arabidopsis*. *J. Exp. Bot.* **65**, 871–884 (2014).

Supplementary Information is available in the online version of the paper.

Acknowledgements This work was supported by funds to N.G. from an ERC Consolidator Grant (GA-N°: 616228 – ENDOFUN), an SNSF grant (31003A_156261), an IEF Marie Curie fellowship (T.G.A.) and an EMBO Long-term postdoctoral fellowship (R.U.). B.D.R., W.S. and B.W. were funded by the Netherlands Organisation for Scientific Research (NWO; VIDI-864.13.001) and The Research Foundation - Flanders (FWO; Odysseus II GOD0515N). We thank A. Paradis and the Central Imaging Facility of the University of Lausanne for support; M. Yamazaki for providing constructs; B. Müller, D. Weijers and T. Vernoux for sharing material; A. Bishopp, A. P. Mähönen, D. Weijers, S. Sabatini, V. Grieneisen, Y. Helariutta and Y. Poirier for discussions; and A. Vjestica, C. Drapek, M. Marek and M. Barberon for input to the manuscript.

Author Contributions T.G.A. planned and conducted all experiments with input from N.G. and J.E.M.V. S.N. conducted initial experiments on PHO1 localization, R.U. created and tested inducible vectors, J.E.M.V. created and tested *shy2-2* lines, and B.D.R., W.S. and B.W. created and selected all *ARR* reporter lines. T.G.A. and N.G. wrote the manuscript. All authors commented on the manuscript.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Publisher's note: Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations. Correspondence and requests for materials should be addressed to N.G. (niko.geldner@unil.ch) or T.G.A. (tonnigrube.andersen@unil.ch).

Reviewer Information Nature thanks H. Fukuda, Y. Helariutta and A. P. Mähönen for their contribution to the peer review of this work.

METHODS

No statistical methods were used to predetermine sample size. The experiments were not randomized and investigators were not blinded to allocation during experiments and outcome assessment.

Plant material and growth conditions. For all experiments, seeds were surface-sterilized in 70% EtOH containing 0.05% Triton-X100, washed twice in 96% EtOH, plated on 1/2 MS (Murashige and Skoog medium) containing 0.8% agar (Duchefa) plates and vernalized at 4°C for 2 days. Seedlings were grown vertically at 22°C, under long-day conditions (18 h, 100 μ E). Unless stated otherwise, all microscopic analyses were performed on roots of 5-day-old seedlings. For hormone and oestradiol (Sigma) treatments, 4-day-old seedlings were transferred to 1/2 MS plates supplemented with hormones or DMSO (mock) for 1 day unless otherwise indicated. Zinc- and iron-deficiency studies were done as previously described². For low phosphate studies, micro-agar containing 10 μ M phosphate (Duchefa Biochemie) and MS without phosphate (Caisson) was used. Plants were grown under 24 h light.

Cloning. The following published mutants and transgenic lines were used in this study: pCASP1::CDEF1²⁵; *ahp6-1*⁶, pAHP6::AHP6-mVenus and pAHP6::AHP6-3mVenus¹⁹; *log3*, *log4* and *log34*, pLOG3::NLS3-GFP and pLOG4::NLS3-GFP⁸; pGPAT5::mCitrine-SYP122²⁸; and pDonor221 containing ARR10_{EAR}²⁶. Reporter constructs for ARR3–10, ARR14–17 and ARR21 were generated using LIC cloning²⁷. pARR::NLS-2YFP plasmids were constructed for the other ARRs, using Gateway technology²⁸ (Life Sciences). The NLS-3mScarlet construct was obtained by DNA synthesis (Thermo Fisher Scientific) into a pDonor221 entry vector. To generate the oestradiol-inducible version of the ELTP promoter, we followed a previously published procedure based on a Gateway-compatible XVE system¹⁸. In brief, 464 bp of the 5'-UTR region of the ELTP (also known as EDA4) gene was cloned into XVE using a KpnI site using Infusion technology (Clontech). A list of primers and promoters can be found in Supplementary Table 1. Corresponding gene numbers listed in supplementary Table 1 or are as follows: *AB11*, At4g26080; *AHP6*, At1g80100; *CASP1*, At2g36100; *CDEF1*, At4g30140; *ELTP* (also known as *EDA4*), At2g48140; *GPAT5*, At3g11430; *LOG3*, At2g37210; *LOG4*, At3g53450; and *SYP122*, At3g52400.

Imaging. Confocal laser scanning microscopy experiments were performed on a Zeiss LSM 880 or a Leica SP8X microscope. All combinatorial fluorescence analyses were run as sequential scans. For fluorescence analysis of marker expression, a Clearsee-based²⁹ protocol was established. Cell walls were stained with calcofluor white (Polysciences) and suberin was stained with Nile red (Sigma). The following settings were used to obtain specific fluorescence signals: EGFP, excitation wavelength (ex): 470 nm, emission wavelength (em): 490–515 nm; mCitrine (with EGFP), ex: 525 nm, em: 530–550 nm; mCitrine (alone), YFP or mVenus, ex: 514 nm, em: 520–550 nm; dsTomato, ex: 561 nm, em: 565–595 nm; mCherry, ex: 594 nm, em: 600–650 nm; Nile red, ex: 561 nm, em: 600–650 nm; and calcofluor white, ex: 405 nm, em: 430–460 nm. fluorol yellow (Sigma) and xylem analysis was done on a Leica DM5000s fluorescence microscope using a GFP filter (ex: 470/40 nm dichroic 500 nm; em: BP 525/50 nm) for fluorol yellow, and a TX2 filter (ex: 560/40 nm dichroic 595 nm; em: 645/75 nm) in combination with differential interference contrast for xylem analysis.

Transcriptional analysis. Total RNA was extracted from 100 mg plant tissue using a Trizol-based PureLink RNA Mini Kit (Thermo Fisher Scientific), then DNase-treated and purified using RNeasy MinElute Cleanup Kit (Qiagen). Reverse transcription was done using a Superscript IV first strand synthesis system (Thermo Fisher Scientific). All steps were performed according to the manufacturers'

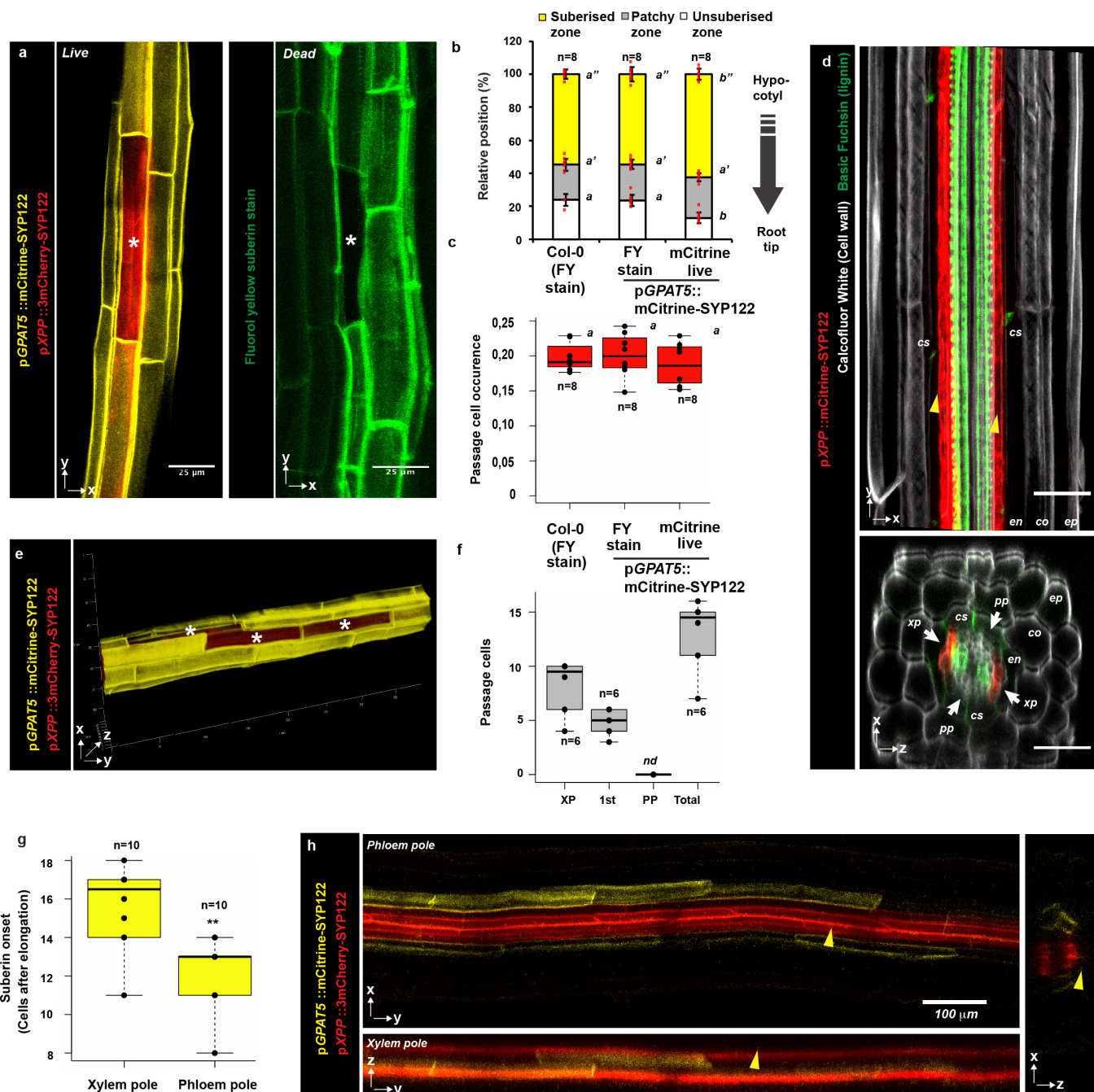
protocols. The PCR reaction was done on a Stratagene Mx3005P thermocycler using a MESA Blue Sybr Green kit. All transcripts are normalized to *UBQ10* expression (see Supplementary Table 1).

Tissue staining and analysis. Unless otherwise noted, suberin lamellae were observed after fluorol-yellow staining as previously described^{2,25,30,31}. Suberin patterns were observed and counted from the hypocotyl junction to the onset of endodermal cell elongation. Three distinct patterns were considered: (1) continuous suberin lamellae; (2) patchy suberin lamellae (corresponding to the area in which individual cells are suberized); and (3) unsuberized cells (corresponding to the youngest part of the root). Passage cells were determined only in the zone containing continuous suberin lamellae. Passage-cell occurrence was obtained by counting the total number of passage cells in both xylem poles, divided by the length (in cells) of the zone. For quantifying the severity of *ahp6-1* xylem defects, the number of xylem-pole-associated endodermal cells above the defective xylem strand were counted, and related to total number of endodermal cells above each strand. Five-day-old roots were cleared using Clearsee²⁹ and stained with basic Fuchsin (Sigma) overnight, as previously reported³².

Statistics and reproducibility. All statistical analyses were done in the R environment³³. For multiple comparisons between genotypes, a one-way ANOVA was performed with a Bonferroni-adjusted ad hoc pairwise two-sided *t*-test. Groups in which differences gave a *P* value lower than 0.05 were considered significantly different. Binary comparisons were performed using a two-tailed Student's *t*-test in Microsoft Excel; *P* values below 0.01 were considered significantly different. All bar graphs represent mean \pm s.d. For all boxplots, the centre depicts the median and the lower and upper box limits depict the 25th and 75th percentile, respectively. Whiskers represent minima and maxima. Closed dots depict individual samples. In cases in which *n* > 10, open dots depict outliers. In all cases, individual biological samples are stated as *n*. All experiments, as well as representative images, were repeated independently at least three times. Individual *P* values for all statistical analyses can be found in Supplementary Table 2.

Data availability. All lines and data generated in this study are available from the corresponding authors upon request.

25. Naseer, S. *et al.* Casparian strip diffusion barrier in *Arabidopsis* is made of a lignin polymer without suberin. *Proc. Natl Acad. Sci. USA* **109**, 10101–10106 (2012).
26. Müller, B. & Sheen, J. Cytokinin and auxin interplay in root stem-cell specification during early embryogenesis. *Nature* **453**, 1094–1097 (2008).
27. Wendrich, J. R., Liao, C.-Y., van den Berg, W. A. M., De Rybel, B. & Weijers, D. Ligation-independent cloning for plant research. *Methods Mol. Biol.* **1284**, 421–431 (2015).
28. Karimi, M., Bleys, A., Vanderhaeghen, R. & Hilson, P. Building blocks for plant gene assembly. *Plant Physiol.* **145**, 1183–1191 (2007).
29. Kurihara, D., Mizuta, Y., Sato, Y. & Higashiyama, T. ClearSee: a rapid optical clearing reagent for whole-plant fluorescence imaging. *Development* **142**, 4168–4179 (2015).
30. Pfister, A. *et al.* A receptor-like kinase mutant with absent endodermal diffusion barrier displays selective nutrient homeostasis defects. *eLife* **3**, e03115 (2014).
31. Roppolo, D. *et al.* A novel protein family mediates Casparian strip formation in the endodermis. *Nature* **473**, 380–383 (2011).
32. Ursache, R., Andersen, T. G., Marhavy, P. & Geldner, N. A protocol for combining fluorescent proteins with histological stains for diverse cell wall components. *Plant J.* (2017).
33. R Development Core Team. *R: A Language And Environment For Statistical Computing* (R Foundation for Statistical Computing, 2015)



Extended Data Figure 1 | Passage cells and suberization in *Arabidopsis*.

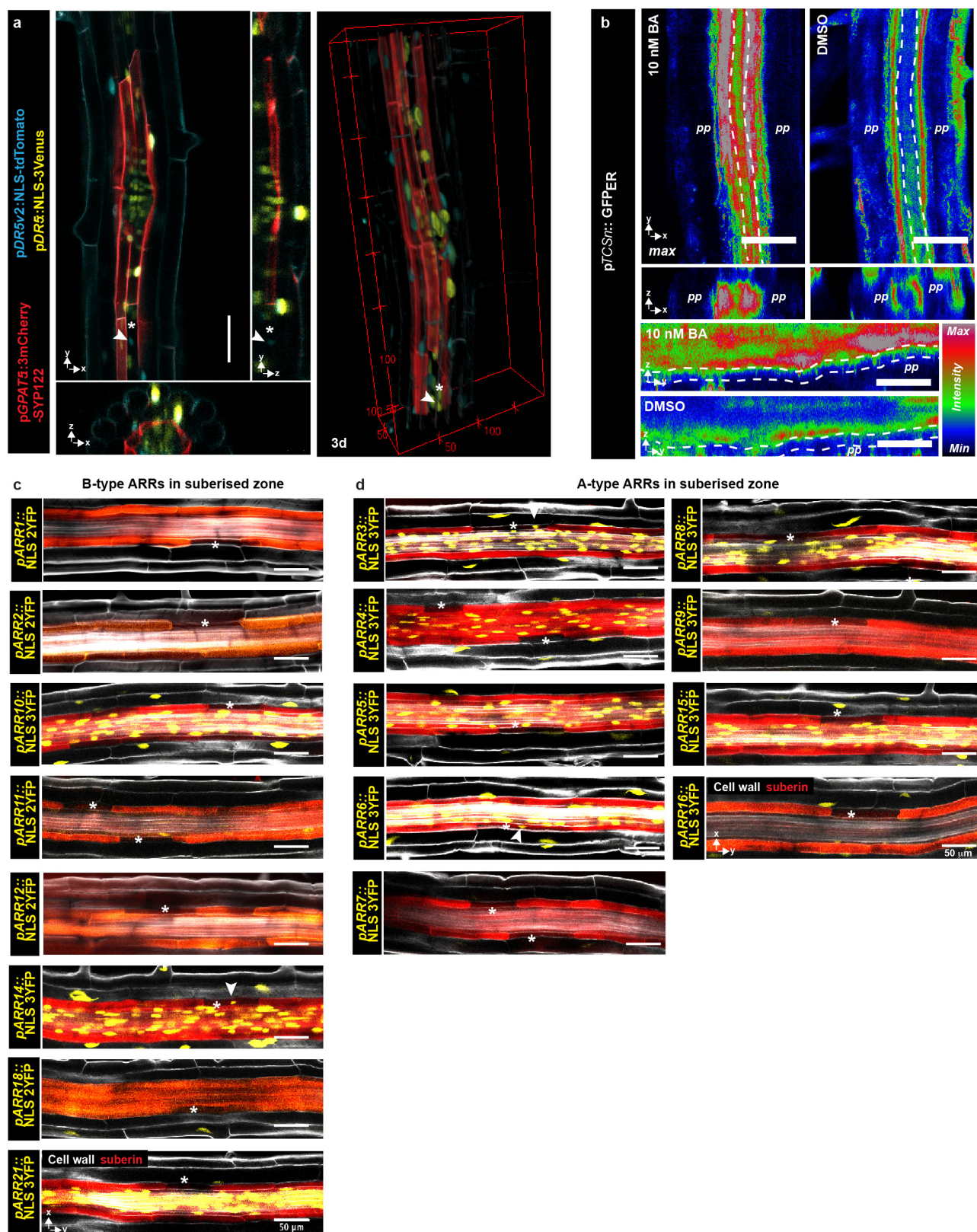
a, A single xylem-pole-associated passage cell surrounded by suberized cells; visualized in live imaging, by expression of the suberin biosynthetic gene *GPAT5* that drives expression of a plasma-membrane-localized mCitrine reporter (mCitrine-SYP122) (yellow, left) or by the suberin-specific dye fluorol yellow (green, right). In the left panel, the xylem-pole pericycle is highlighted using the promoter pXPP driving a 3mCherry reporter fused to the plasma-membrane-localized SYP122 (3mCherry-SYP122). Note that the fluorol-yellow protocol requires heating the sample to 70 °C, and is incompatible with fluorescent protein detection.

b, Comparison of endodermal suberization in Col-0 and in plants expressing the *GPAT5*-based reporter, determined either by the suberin-specific dye fluorol yellow or live mCitrine expression, respectively. Red dots, individual data points.

c, Comparison of passage-cell occurrence in Col-0 and in plants expressing the *GPAT5*-based reporter, determined either by the suberin-specific dye fluorol yellow or live mCitrine expression. Dots, individual data points.

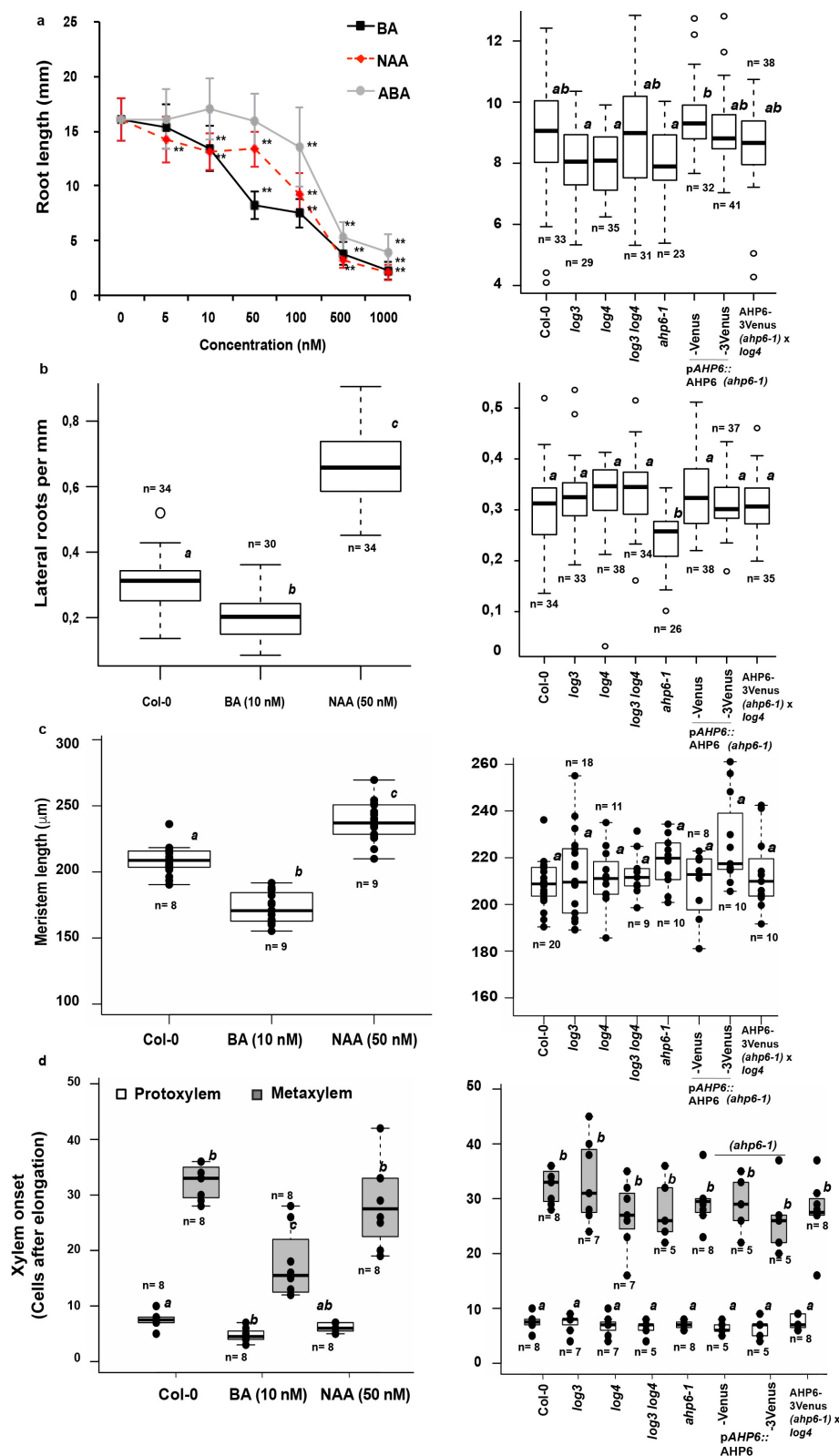
d, Expression of a plasma-membrane-localized mCitrine-SYP122 marker (red) driven by the xylem-pole-specific promoter pXPP, in the zone of protoxylem onset. Roots were cleared using ClearSee and stained with basic fuchsin (green) and

calcofluor white (grey)³². **e**, Radial- and longitudinally connected passage cells in the xylem pole of the suberized zone. **f**, Direct quantification of passage cells residing directly above xylem pole (XP), first side cell to the xylem pole (1st) or phloem pole (PP) in five-day-old Col-0 plants. Dots, individual data points. **g**, Quantification of the onset of endodermal suberization in the phloem and xylem poles of roots expressing markers as in **a**. Dots, individual data points. **h**, Representative images of the patchy zone of endodermal suberization in the phloem and xylem poles of roots used for quantification in **f**. Yellow arrowheads, xylem pole; asterisks, passage cells. cs; Casparian strip, FY; fluorol yellow. For all stacked graphs, there are three measurements per root: unsuberized zone, white; patchy zone, grey; and suberized zone, yellow. Bar graphs represent mean \pm s.d. and boxplots show median. Statistically significant differences between groups were tested using a post hoc Bonferroni-adjusted paired two-sided *t*-test. For more information on data plots, see Methods. **a** and **e** represent five independent lines; all stainings were repeated three times *n*, independent biological samples. ***P* < 0.01, two-tailed *t*-test. For individual *P* values, see Supplementary Table 2. Scale bars, 25 μ m, unless otherwise specified.



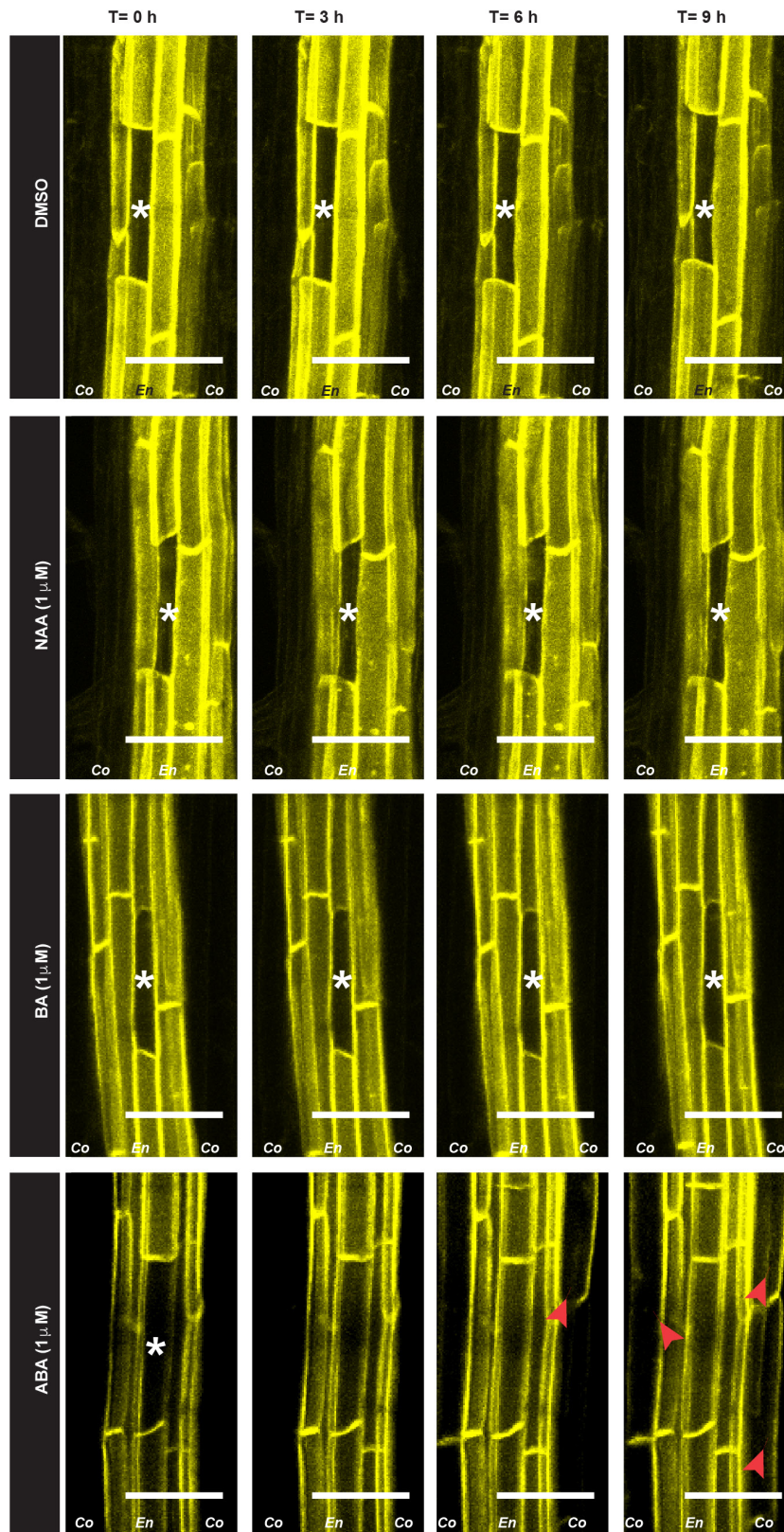
Extended Data Figure 2 | Auxin and cytokinin signalling in the suberized root zone. **a**, Activity of the auxin-signalling reporter DR5 (yellow) as well as the highly sensitive DR5v2 (blue) in an area of the suberized zone of a five-day-old root with an emerging lateral root. Suberized cells were visualized based on the suberin biosynthetic gene *GPAT5* driving expression of a plasma-membrane-localized 3mCherry-based reporter. **b**, Expression of endoplasmic-reticulum-localized GFP driven by the cytokinin signalling reporter *TCSn* in the phloem and xylem poles of 5-day-old roots in the suberized zone. Plants were either

grown on plates containing 5 nM cytokinin (BA) or a mock treatment (DMSO). Punctured lines indicate the endodermis. **c**, Expression of B-type ARRs in the suberized endodermis of five-day-old roots, suberin and cellwalls are stained using a Clearsee protocol in combination with Nile red and calcofluor white, respectively³². **d**, Expression of A-type ARRs in the suberized endodermis of five-day-old roots, suberin and cell walls, as in **c**. All stainings were repeated three times. In **a** and **b**, images are representative of eight independent lines. White arrowheads, passage-cell nuclei. Asterisks, passage cells. Scale bars, 50 μ m.



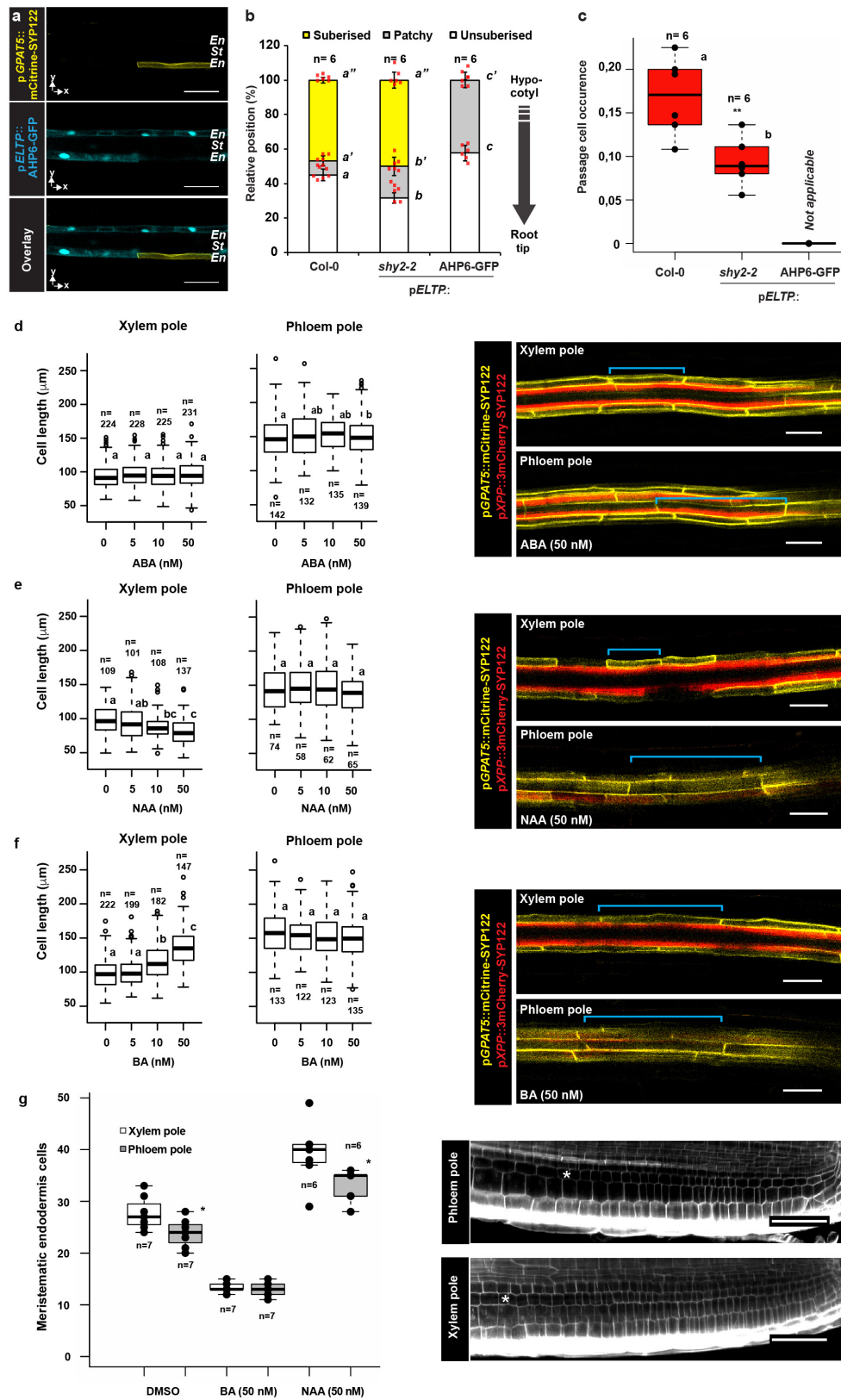
Extended Data Figure 3 | Phenotypic analysis of hormone-treated plants and mutants. **a**, Root length of five-day-old plants either germinated on increasing concentration of cytokinin (BA), auxin (NAA) or abscisic acid (ABA) (left), or of the mutants used in this study (right). Open circles, outliers. **b**, Emerged lateral roots per mm of ten-day-old Col-0 roots germinated in presence of DMSO, 10 nM cytokinin (BA) or 50 nM auxin (NAA) (left), or of the mutants used in this study (right). Open circles, outliers. **c**, Length of the root apical meristem of five-day-old Col-0 plants germinated on either DMSO, 10 nM cytokinin (BA) or 50 nM auxin (NAA) (left), or of mutants used in this study (right). Dots,

individual data points. **d**, Onset of protoxylem and metaxylem after the elongation zone of five-day-old seedling germinated in presence of DMSO, 10 nM cytokinin (BA) or 50 nM auxin (NAA) (left), or of mutants used in this study (right). Dots, individual data points. Bar graphs represent mean \pm s.d. and boxplot centres show median. Statistically significant differences between groups were tested using a post hoc Bonferroni-adjusted paired two-sided *t*-test. For more information on data plots, see Methods. *n*, independent biological samples. In **a**, $**P < 0.01$, two-tailed *t*-test, *n* = 25 independent biological replicas for each treatment.



Extended Data Figure 4 | Hormone-induced closure of passage cells. Time-course analysis showing behaviour of passage cells, visualized by mCitrine-SYP122 driven by the promoter of the suberin synthesis marker gene *GPAT5* for 9 h on 1/2 MS medium containing DMSO, cytokinin (BA),

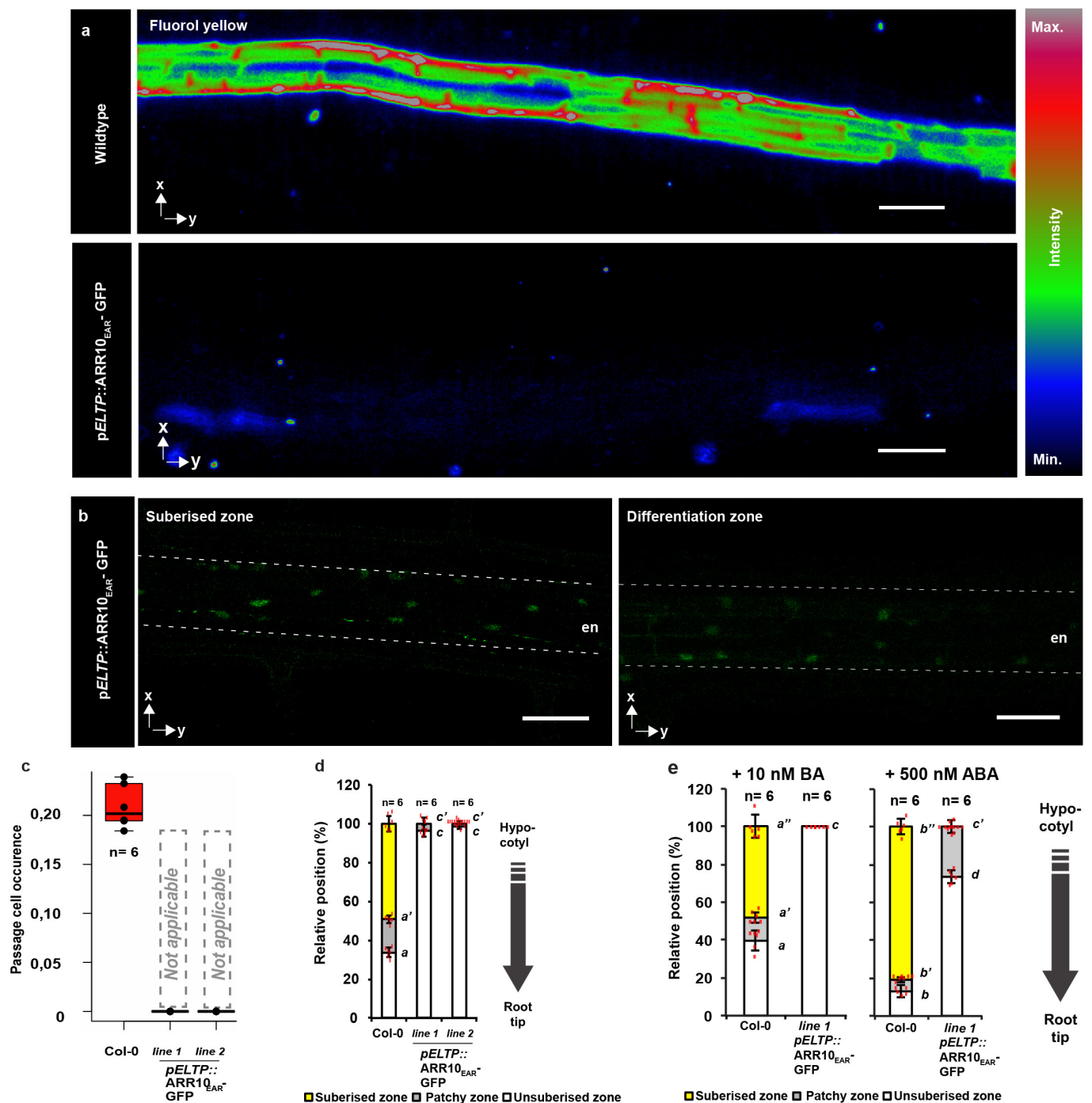
auxin (NAA) or abscisic acid (ABA). Red arrowheads point to ectopic suberization in the cortex on ABA treatment. Asterisks, passage cells. All time courses were repeated three times. Scale bars, 50 μm.



Extended Data Figure 5 | See next page for caption.

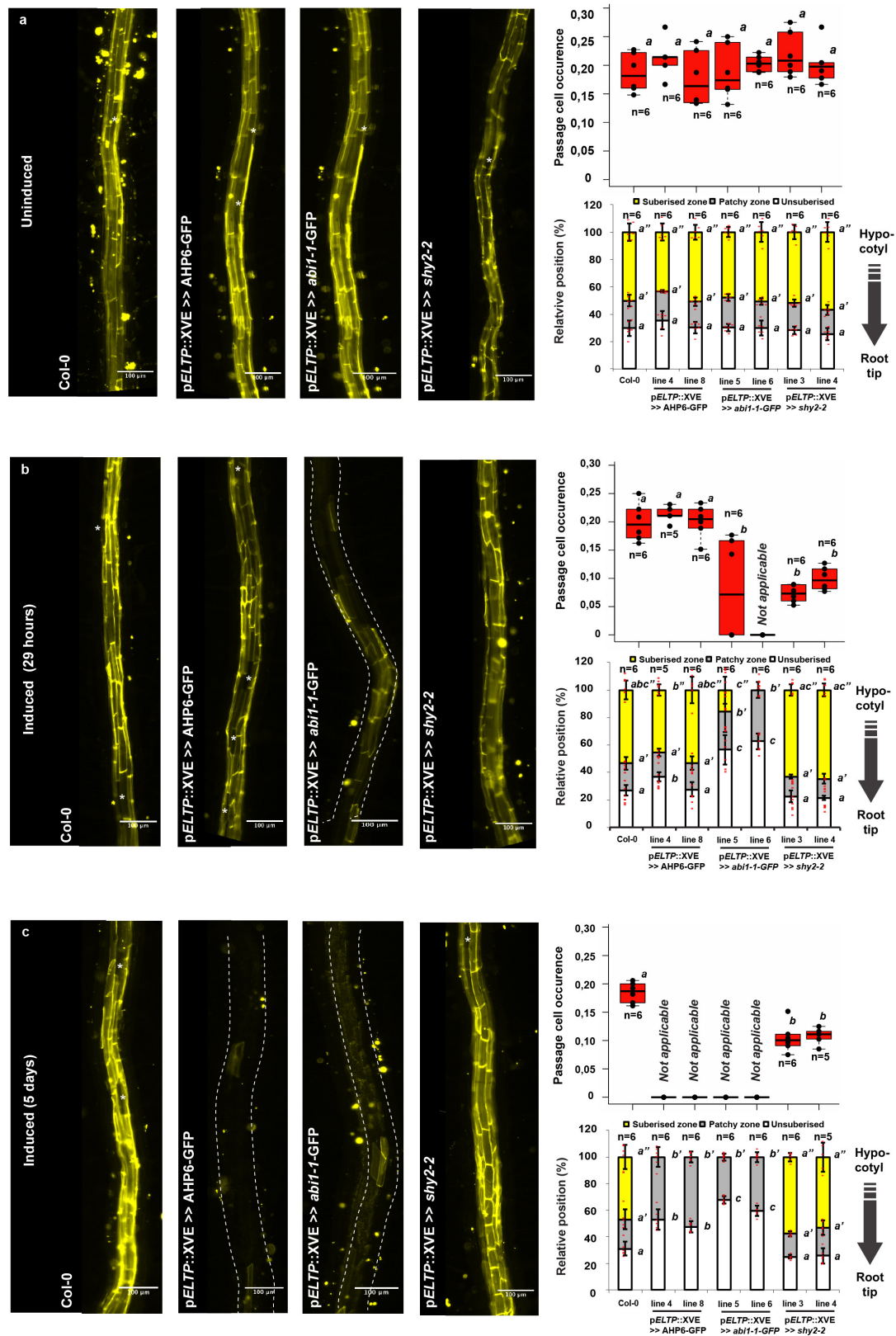
Extended Data Figure 5 | Hormone-induced changes in endodermal cell lengths. **a**, Expression of mCitrine-SYP122 driven by the *GPAT5* promoter (yellow), in the fully suberized zone in plants that express the dominant cytokinin-signalling inhibitor AHP6 fused to GFP in the differentiated endodermis using the *ELTP* promoter (blue). **b**, Endodermal suberization in five-day-old seedlings that express AHP6-GFP or the dominant auxin signalling repressor *shy2-2* in the differentiated endodermis. Red dots, individual data points. **c**, Occurrence of passage cells in five-day-old seedlings that express AHP6-GFP or the auxin signalling repressor *shy2-2* in the differentiated endodermis. Dots, individual data points. **d–f**, Length of suberized endodermal cells measured in the suberized zone of plants expressing pGPAT5::mCitrine-SYP122 and pXPP::3mCherry-SYP122 constructs. Open circles, outliers. Plants were grown for five days on 1/2 MS medium containing depicted concentrations of abscisic acid (**d**), auxin and naphthalene acetic acid (**e**), and cytokinin and benzyl adenine (**f**).

g, The number of meristematic cells in the phloem- and xylem-pole endodermis was counted in cleared roots stained with calcofluor white³² to highlight cell walls. Dots, individual data points. For all stacked graphs, there are three measurements per root: unsuberized zone, white; patchy zone, grey; and suberized zone, yellow. Protoxylem or -phloem was used as a marker to identify the poles. Cells were counted from the quiescent centre to the onset of elongation. Asterisks, last cell of the division zone. Blue lines, lengths of individual cells. Bar graphs represent mean \pm s.d. and boxplot centres show median. Statistically significant differences between groups were tested using a post hoc Bonferroni-adjusted paired two-sided *t*-test. For more information on data plots, see Methods. In **a**, images are representative of 12 independent lines. In **g**, all stainings were repeated three times. *n*, number of measurements across 16 independent biological samples. ***P* < 0.01, two-tailed *t*-test. Scale bars, 25 μ m.



Extended Data Figure 6 | Manipulation of suberization through signalling repression. **a**, Representative pictures of fluorol-yellow staining of suberin in lines expressing $ARR10_{EAR}-GFP$ in all differentiated endodermal cells, driven by the $ELTP$ or $CASP1$ promoters. **b**, Occurrence of passage cells in lines expressing $ARR10_{EAR}-GFP$ in all differentiated endodermal cells, using the $ELTP$ and $CASP1$ promoters. **c**, Behaviour of suberin in lines expressing $ARR10_{EAR}-GFP$ in all differentiated endodermal cells, using the $ELTP$ and $CASP1$ promoters. **d**, Behaviour of suberin in five-day-old plants expressing $ARR10_{EAR}-GFP$ in all differentiated endodermal cells, using the $ELTP$ and $CASP1$ promoters. Red dots, individual data points. **e**, Behaviour of suberin in five-day-old

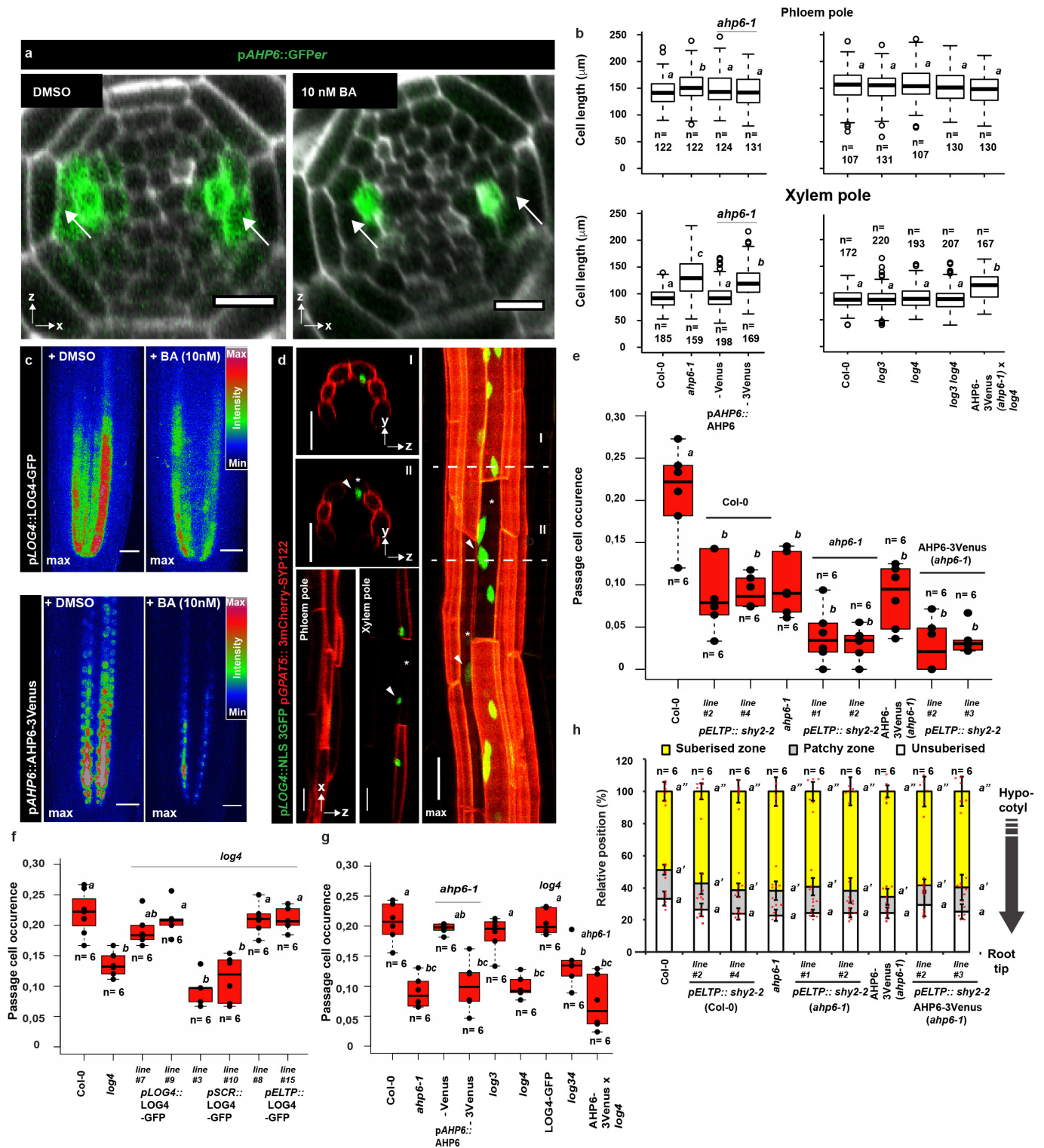
plants expressing $ARR10_{EAR}-GFP$ in all differentiated endodermal cells, using the $ELTP$ and $CASP1$ promoters, when plants were germinated on either cytokinin (BA) or abscisic acid (ABA). Red dots, individual data points. For all stacked graphs, there are three measurements per root: unsuberized zone, white; patchy zone, grey; and suberized zone, yellow. Bar graphs represent mean \pm s.d. and boxplots show median. Statistically significant differences between groups were tested using a post hoc Bonferroni-adjusted paired two-sided t -test. For more information on data plots, see Methods. In **b**, images are representative of nine independent lines. All stainings were repeated three times. n , independent biological samples. Scale bars, 50 μm .



Extended Data Figure 7 | Temporal control of suberin inhibition.

Expression of temporally controlled dominant inhibitors of cytokinin (AHP6-GFP), ABA (*abi1-1-GFP*) or auxin (*shy2-2*) signalling in the differentiated endodermis, using an oestradiol-inducible system (XVE)¹⁹. **a–c**, Fluorol-yellow staining (left) and quantification of passage cells (top right) and suberin (bottom right) in the late endodermis of plants grown on 1/2 MS medium containing DMSO for five days (**a**), plants grown for four days on 1/2 MS medium containing DMSO, then transferred to 5 μ M oestradiol for 29 h (**b**) or germinated for five days on 1/2-MS-medium plates containing 5 μ M oestradiol (**c**). Asterisks, passage cells. Scale bars, 100 μ m.

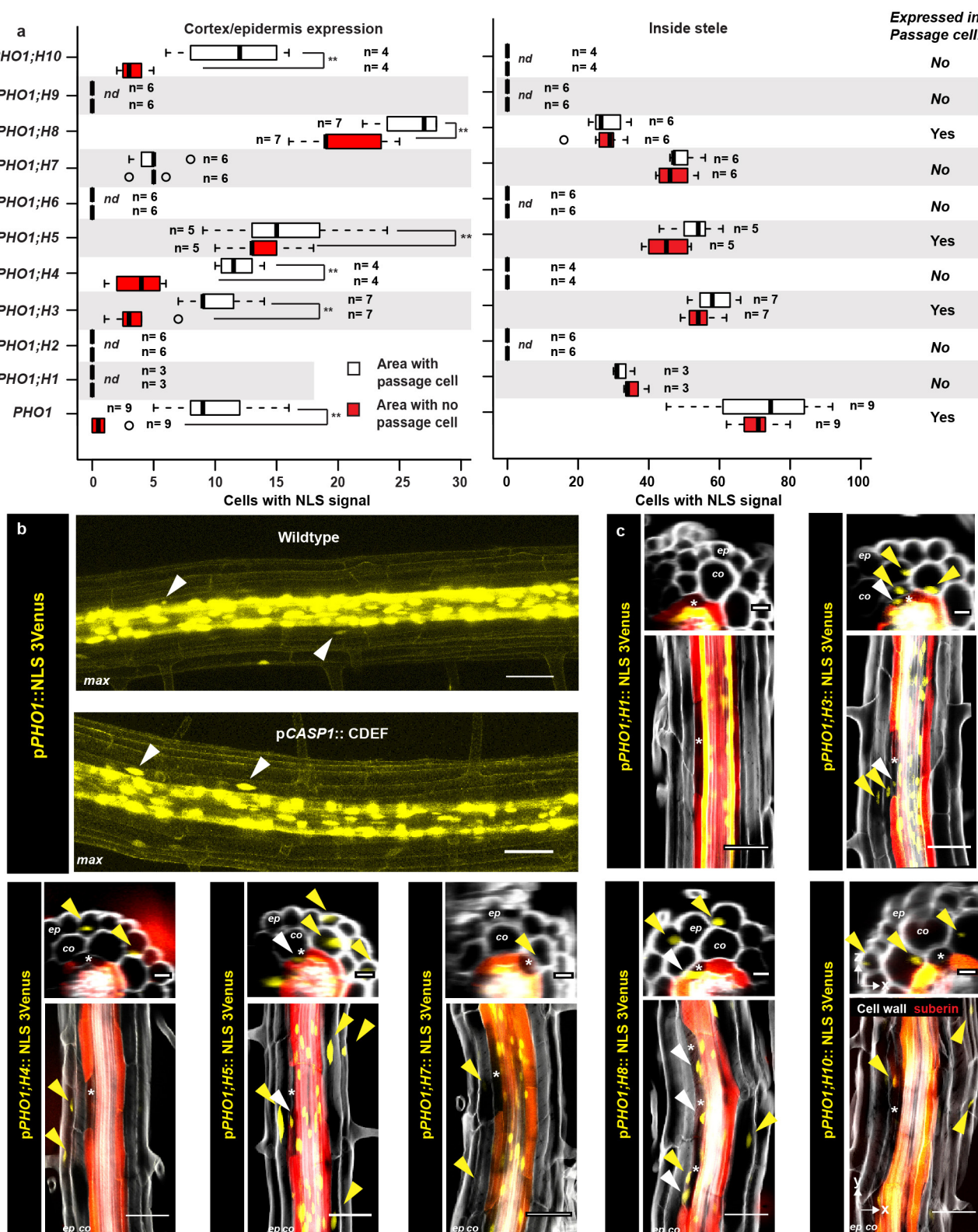
All dots, individual data points. For all stacked graphs, there are three measurements per root: unsuberized zone, white; patchy zone, grey; and suberized zone, yellow. *n* represent biologically independent samples. Bar graphs represent mean \pm s.d. and boxplot centres show median. Statistically significant differences between groups were tested using a post hoc Bonferroni-adjusted paired two-sided *t*-test. For more information on data plots, see Methods. In **a–c**, images are representative of nine independent lines. All stainings were repeated three times. *n*, independent biological samples. Scale bars, 100 μ m.



Extended Data Figure 8 | See next page for caption.

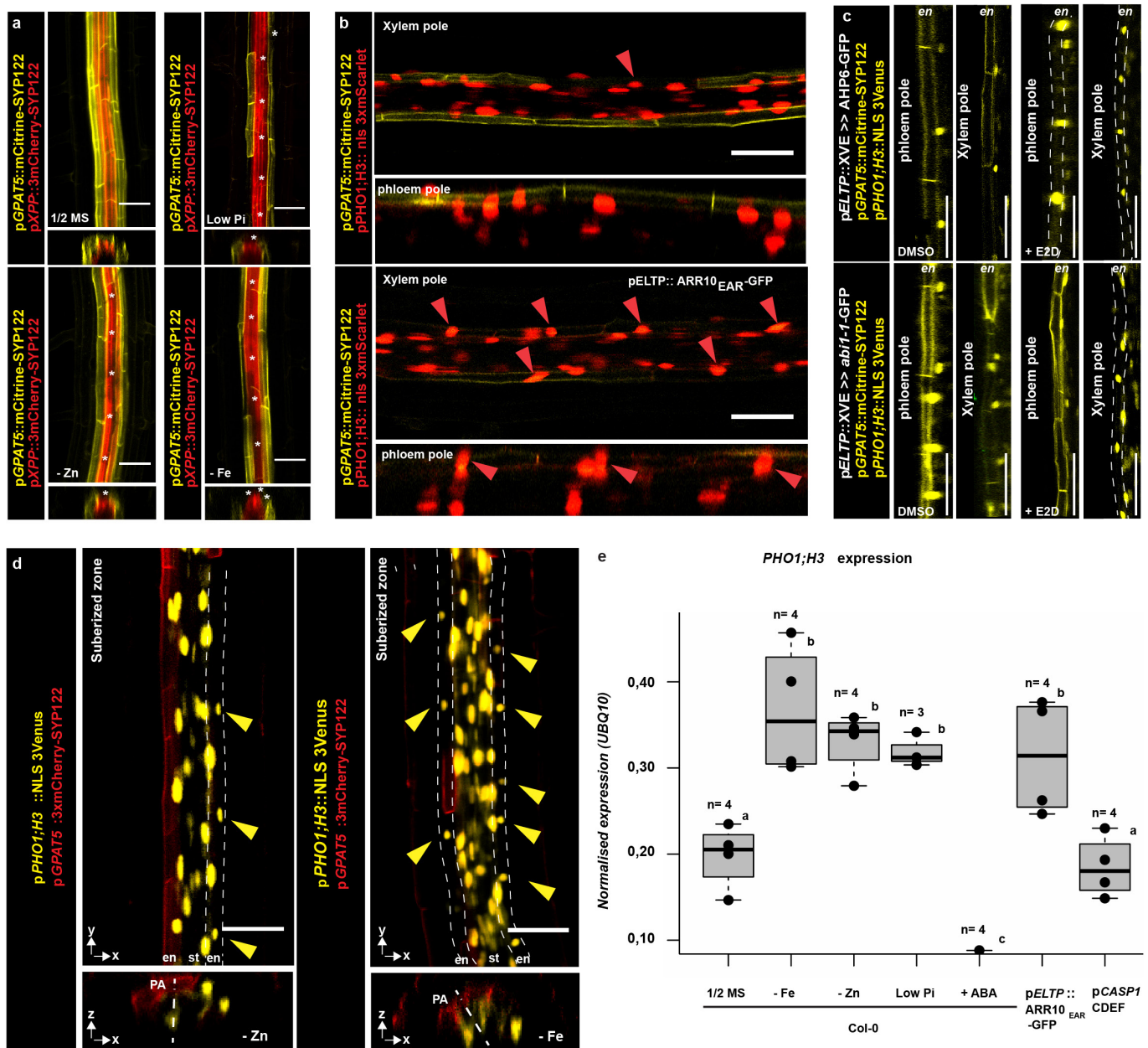
Extended Data Figure 8 | Cytokinin-dependent formation of passage cells. **a**, Expression of *AHP6* visualized by endoplasmic-reticulum-localized GFP (GFP_{ER})¹⁹, in the root apical meristem of plants grown either in presence of a mock treatment (DMSO) or 10 nM cytokinin (BA). Arrows point to xylem-pole pericycle cells. **b**, Length of suberized endodermal cells in the phloem and xylem poles of five-day-old *ahp6-1* and *log* mutants, complementing lines and stacked mutants grown on 1/2-MS-medium plates. Data partly overlap with those in Fig. 3. Open circles, outliers. **c**, Expression of LOG4–GFP and AHP6–3mVenus driven by their native promoters, in the root apical meristem of plants grown either in presence of mock treatment (DMSO) or 10 nM cytokinin (BA). **d**, Expression of *LOG4* in the suberized zone of five-day-old seedlings visualized by the native *LOG4* promoter driving expression of an NLS–3GFP construct. Suberized endodermal cells were highlighted by expression of the suberin biosynthetic marker pGPAT5::3mCherry-SYP122. **e**, Occurrence of passage cells in lines with inhibited AHP6 diffusion in combination with repressed auxin signalling (through the expression of *shy2-2*) in the differentiated endodermis. Dots, individual data points. **f**, Passage-cell occurrence in the *log4* mutant complemented

with a LOG4–GFP construct driven by either the native promoter (*pLOG4*), an early endodermis promoter (*pSCR*) or a differentiating endodermis promoter (*pELTP*). Dots, individual data points. **g**, Occurrence of passage cells in the suberized zone of five-day-old *ahp6-1*, *log* mutants and *ahp6 log* additionally complemented by non-mobile AHP6 to obtain endodermis-specific defects. Dots, individual data points. **h**, Behaviour of suberin in plants with or without inhibited auxin perception in the late endodermis grown on 1/2-MS-medium plates. Red dots, individual data points. For all stacked graphs, there are three measurements per root: unsuberized zone, white; patchy zone, grey; and suberized zone, yellow. Bar graphs represent mean \pm s.d. and boxplot centres show median. Individual letters show significantly different groups according to a post hoc Bonferroni-adjusted paired two-sided *t*-test. For more information on data plots, see Methods. In **c** and **d**, images represent eight independent lines. All stainings were repeated three times. *n*, independent biological samples. In **b**, *n* represents measurements across 16 biologically independent samples. Scale bars, 10 μ m (**a**) or 25 μ m (**c**, **d**).



Extended Data Figure 9 | The PHO1 transporter gene family as marker for passage cells. **a**, Expression of PHO1 family members in the cortex and/or epidermis (left) and inside the stele (right) within the suberized zone of five-day-old roots grown on 1/2 MS medium. Individual nuclei showing expression of the respective NLS reporter were counted in similar 3D stacks, taken either in a fully suberized area (red bars) or in an area containing one passage cell (white bars). Open circles, outliers. **b**, Expression of *PHO1* in five-day-old seedlings in the late endodermis of wild type, or a suberin-degradation line with virtually absent suberin (pCASP1::CDEF1). **c**, Representative images of the expression of each member of the PHO1 family in roots of five-day-old seedlings,

highlighted by the NLS-3mVenus reporter. Seedlings were stained with Nile red for suberin and calcofluor white for cell walls. Scale bars in x - z projections represent 10 μm . White arrowheads, passage-cell nuclei; yellow arrowheads point towards nuclei of cortex and/or epidermal cells. Asterisks, passage cells. n.d., not detected. Boxplot centres show median. $**P < 0.01$, two-sided t -test. For more information on data plots, see Methods. In **b** and **c**, images represent 15 and 11 independent lines, respectively. All stainings were repeated three times. n , independent biological samples. Unless otherwise stated, scale bars represent 25 μm . For insets, scale bars represent 10 μm .



Extended Data Figure 10 | Nutrient deficiency responses of suberization in the xylem pole. **a**, Expression of the suberization marker *GPAT5* and the xylem-pole-pericycle marker *XPP* using a mCitrine-SYP122 or a 3mCherry-SYP122 plasma-membrane-anchored reporter, respectively. Plants were germinated for five days on 1/2 MS medium (1/2 MS), plates with low (10 μ M) inorganic phosphate (low Pi), no zinc (-Zn) or no iron (-Fe). Images representative of four independent lines. **b**, Expression of the suberization marker *GPAT5* and *PHO1;H3* using a mCitrine-SYP122 or a NLS-3mScarlet reporter, respectively, in the suberized zone of Col-0 plants or plants expressing the dominant cytokinin signalling repressor *ARR10-GFP*, in all differentiating endodermal cells. Each image is representative of eight independent lines. **c**, Expression of the suberization marker *GPAT5* and *PHO1;H3* using a mCitrine-SYP122 or a NLS-3mVenus reporter, respectively, in the suberized zone of plants expressing oestradiol-inducible (XVE) versions of the dominant cytokinin signalling repressor *AHP6-GFP* or the dominant ABA signalling repressor *abi1-1*, in all differentiated endodermal cells. Plants were grown for five days in presence of DMSO or 5 μ M oestradiol (+E2D), all treatments were repeated three times. Each image is

representative of five independent lines. **d**, Expression of *PHO1;H3* and the suberization marker *GPAT5* in roots of five-day-old Col-0 plants grown on 1/2 MS medium, plates containing no iron, no zinc or with low (10 μ M) inorganic phosphate. Each image is representative of nine independent lines. All treatments were repeated three times. **e**, Transcriptional analysis of *PHO1;H3* expression in roots of five-day-old Col-0 plants grown on 1/2 MS medium, plates containing no iron, no zinc or with low (10 μ M) inorganic phosphate, in plants expressing the dominant cytokinin signalling repressor *ARR10-GFP* in all differentiated endodermal cells, or in plants with strongly reduced suberin deposition by expression of a suberin-degrading enzyme (*pCASP1::CDEF*). Expression was normalized to that of *UBQ10*. Dots, individual data points. For all stacked graphs, there are three measurements per root: unsuberized zone, white; patchy zone, grey; and suberized zone, yellow. Asterisks highlight unsuberized xylem-pole endodermal cells. Boxplot centres show median. Statistically significant differences between groups were tested using a post hoc Bonferroni-adjusted paired two-sided *t*-test. For more information on data plots, see Methods. *n*, independent biological samples. Scale bars, 25 μ m.

Carbonate-sensitive phytotransferrin controls high-affinity iron uptake in diatoms

Jeffrey B. McQuaid^{1,2}, Adam B. Kustka³, Miroslav Oborník^{4,5}, Aleš Horák^{4,5}, John P. McCrow¹, Bogumil J. Karas^{1,†}, Hong Zheng¹, Theodor Kindeberg², Andreas J. Andersson², Katherine A. Barbeau² & Andrew E. Allen^{1,2}

In vast areas of the ocean, the scarcity of iron controls the growth and productivity of phytoplankton^{1,2}. Although most dissolved iron in the marine environment is complexed with organic molecules³, picomolar amounts of labile inorganic iron species (labile iron) are maintained within the euphotic zone⁴ and serve as an important source of iron for eukaryotic phytoplankton and particularly for diatoms⁵. Genome-enabled studies of labile iron utilization by diatoms have previously revealed novel iron-responsive transcripts^{6,7}, including the ferric iron-concentrating protein ISIP2A⁸, but the mechanism behind the acquisition of picomolar labile iron remains unknown. Here we show that ISIP2A is a phytotransferrin that independently and convergently evolved carbonate ion-coordinated ferric iron binding. Deletion of *ISIP2A* disrupts high-affinity iron uptake in the diatom *Phaeodactylum tricornutum*, and uptake is restored by complementation with human transferrin. ISIP2A is internalized by endocytosis, and manipulation of the seawater carbonic acid system reveals a second-order dependence on the concentrations of labile iron and carbonate ions. In *P. tricornutum*, the synergistic interaction of labile iron and carbonate ions occurs at environmentally relevant concentrations, revealing that carbonate availability co-limits iron uptake. Phytotransferrin sequences have a broad taxonomic distribution⁸ and are abundant in marine environmental genomic datasets^{9,10}, suggesting that acidification-driven declines in the concentration of seawater carbonate ions will have a negative effect on this globally important eukaryotic iron acquisition mechanism.

The iron cargo protein transferrin is thought to have originated in metazoa as an adaptation to multicellularity¹¹. In vertebrates, serum transferrin circulates in the blood, binding labile ferric iron (Fe³⁺) with high affinity ($K_d^{\text{Fe}^{3+}} \approx 1\text{--}10\text{ pM}$) and distributing iron to cells via receptor-mediated endocytosis¹². Transferrin binds iron tightly by coordinating ferric iron with an exogenous carbonate anion (CO₃²⁻). Following endocytosis, the carbonate anion is protonated and ferric iron is reduced, disrupting Fe coordination and initiating the release of iron¹³. Recent discoveries of ferric-iron-concentrating ISIP2A proteins in diatoms^{7,8} and transferrin in the green alga *Dunaliella salina*¹⁴ suggest that transferrin-like iron acquisition mechanisms are not limited to multicellular organisms. We used a marine microeukaryote transcriptional dataset to reconstruct the phylogenetic histories of ISIP2A and transferrin, revealing a common origin in phosphonate-binding periplasmic proteins (PBPs; Fig. 1). Transferrins and the ISIP2A clade of algal transferrin-like proteins (herein phytotransferrin, or pTF¹⁵) are functional analogues that have convergently added iron coordination sites onto an anion-binding precursor protein (Extended Data Fig. 1a). The evolution of phytotransferrin marks the third occasion in which anion-coordinated iron binding has arisen from the bacterial PBP family¹⁶, highlighting the fundamental role coordinating anions have within the transferrin superfamily¹³.

Molecular dating (by PhyloBayes) estimates the appearance of phytotransferrins to between 914 and 671 million years ago, which is consistent with the inferred changes to Neoproterozoic marine redox states¹⁷ and the subsequent requirements to develop mechanisms to access dilute ferric iron (Extended Data Fig. 2). Transferrin and phytotransferrin are simultaneously present in the genomes of deeply branching chlorophyte algae (Extended Data Fig. 3), although our phylogenetic reconstruction suggests that these proteins were acquired through separate events. Whereas transferrin appears to have been vertically inherited from marine archaea, the chlorophyte acquisition of phytotransferrin is less clear, plausibly occurring through endosymbiotic or horizontal gene transfer. Unlike transferrins, phytotransferrins are often transmembrane anchored (Supplementary Table 2), a selective advantage for unicellular organisms that probably contributes to its wide phylogenetic distribution among marine plankton.

To examine the role of phytotransferrin in high-affinity iron uptake in diatoms, we used TALE nucleases to disrupt *ISIP2A* (hereafter Δ ISIP2A) in *P. tricornutum* (Fig. 2a and Extended Data Fig. 4a). In artificial seawater in which the concentration of labile inorganic iron species ([Fe']) is controlled with excess EDTA, Δ ISIP2A has significantly reduced growth rates at low [Fe'] (Extended Data Fig. 4b). Short-term ⁵⁹Fe uptake rates of Δ ISIP2A are reduced by more than 90%, confirming the role of ISIP2A in high-affinity iron uptake (Fig. 2b). In *P. tricornutum*, ISIP2A is one of three transmembrane-anchored phytotransferrin homologues (Extended Data Fig. 1b); however, ISIP2A dominates transcriptional abundance, accounting for 95% of all phytotransferrin transcripts¹⁸. Uptake of ⁵⁹Fe-ferrioxamine B is not affected in Δ ISIP2A, which suggests a separate mechanism for the uptake of complexed iron (Extended Data Fig. 4c).

To determine whether ISIP2A is a functional analogue of transferrin, codon-optimized N- and C-terminal domains of human serum transferrin (hTF) were fused to the signal peptide and transmembrane anchor of ISIP2A and reintroduced into Δ ISIP2A via conjugation¹⁹. The N-terminal domain of hTF fully restored high-affinity iron uptake equal to that of ISIP2A, whereas the C-domain partially restored uptake (Fig. 2b), confirming phytotransferrin to be a functional analogue of transferrin and suggesting that both proteins use similar iron-binding, -internalization and -release mechanisms. In hTF, mutation of the conserved tyrosine ligands disrupts iron coordination¹¹. In ISIP2A, mutagenesis of the analogous tyrosines decreases short-term uptake rates by 50%, providing further evidence of a shared mechanism (Fig. 2c).

Internalization through endocytosis is required for the release of transferrin-bound iron²⁰. In *P. tricornutum*, inclusion of an endocytosis inhibitor decreases short-term ⁵⁹Fe' uptake (Extended Data Fig. 4d). When ISIP2A is labelled with fluorescent protein, ISIP2A-RFP localizes to the outer membrane, small internal vesicles and endosome-like

¹J. Craig Venter Institute, Microbial and Environmental Genomics, La Jolla, California 92037, USA. ²Scripps Institution of Oceanography, University of California, La Jolla, California 92093, USA. ³Rutgers University–Newark, Earth and Environmental Sciences, Newark, New Jersey 07102, USA. ⁴Biology Centre CAS, Institute of Parasitology, Branišovská 31, 370 05 České Budějovice, Czech Republic. ⁵University of South Bohemia, Faculty of Science, Branišovská 31, 370 05 České Budějovice, Czech Republic. [†]Present address: Designer Microbes Inc., London, Ontario N6G4X8, Canada.

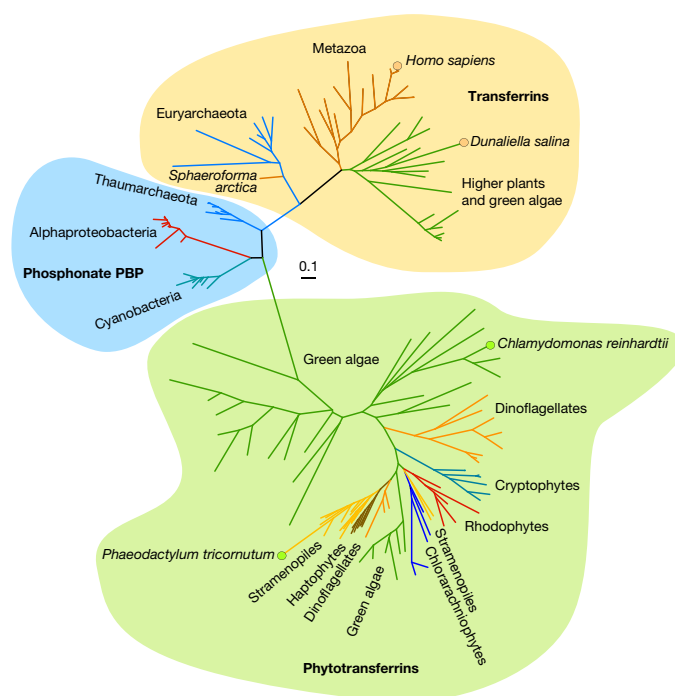


Figure 1 | Transferrin and phytotransferrin are functional analogues with a common origin. Maximum likelihood phylogenetic tree of the divergence of transferrin and phytotransferrin from ancestral PBPs as inferred from amino acid sequences. Branch colour denotes phylogenetic group, and circles at branch ends denote characterized iron-binding proteins: hTF (*H. sapiens*), triplicated transferrin (TTF1, *D. salina*), Fe-assimilation protein (FEA1, *Chlamydomonas reinhardtii*) and ISIP2A (*P. tricornutum*). Scale bar, 0.1 amino acid substitutions per position.

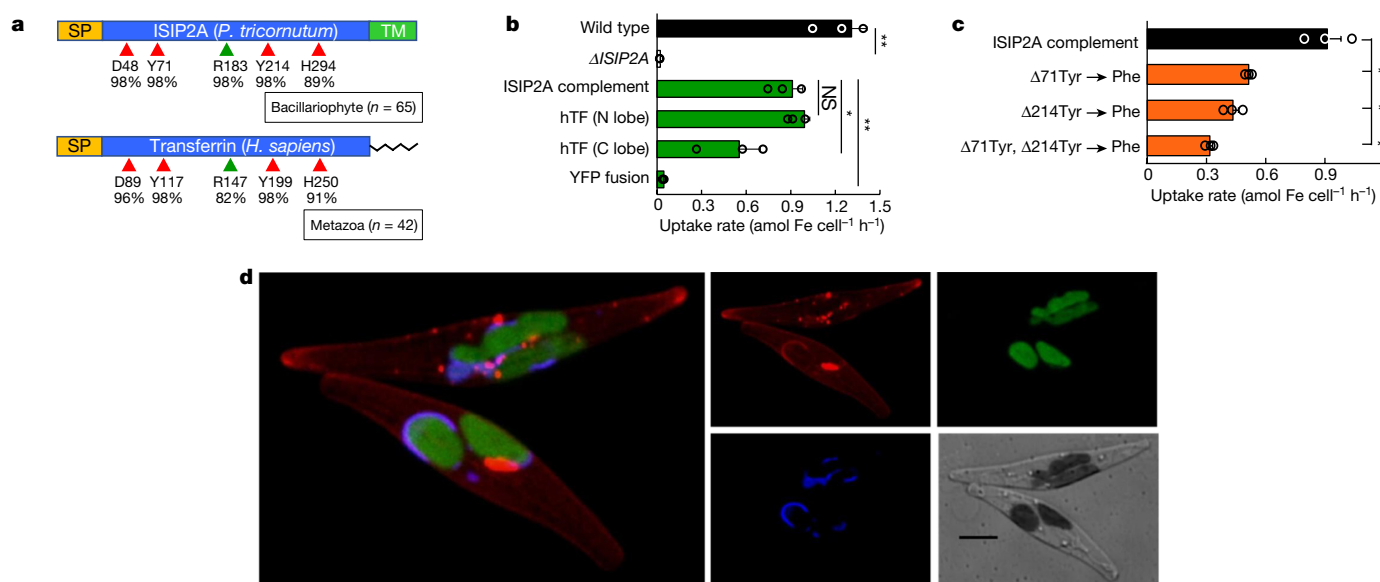


Figure 2 | Characterization of phytotransferrin ISIP2A (*P. tricornutum*). **a**, Schematic of phytotransferrin (ISIP2A) and hTF (N-terminal domain). Locations and per cent conservation of putative iron- and carbonate-coordinating amino acids (red and green triangles, respectively). SP, signal peptide; TM, transmembrane domain. **b**, High-affinity iron uptake (in attomole (amol) Fe cell⁻¹ h⁻¹) is disrupted in Δ ISIP2A and restored in hTF- and ISIP2A-complemented cultures. YFP, yellow fluorescent

structures, a pattern that is consistent with internalization through endocytosis²⁰ (Fig. 2d). The receptor-like iron-regulated protein ISIP1^{6,7} was co-visualized, however cell-surface interactions with ISIP2A were not apparent (Fig. 2d). When iron-limited *P. tricornutum* is stained with a membrane-impermeable dye and pulsed with Fe³⁺, portions of the outer membrane are internalized (Extended Data Fig. 4e). Because endocytosis is restricted to eukaryotic organisms, phytotransferrin is a factor that probably contributes to the rapid response of diatoms and other eukaryotic plankton to iron fertilization events in high-nutrient low-chlorophyll regions².

In both human¹² and algal²¹ transferrin, CO₃²⁻ is required for the efficient binding of iron. To investigate the role of CO₃²⁻ in phytotransferrin-mediated Fe³⁺ uptake, we resuspended cultures of *P. tricornutum* that were acclimatized to a low-iron environment into pH-controlled ⁵⁹Fe-uptake medium and manipulated carbonate ion availability [CO₃²⁻] by NaHCO₃ addition (Extended Data Table 1). At both low and high [Fe³⁺], short-term Fe³⁺ uptake rates were positively correlated with [CO₃²⁻], consistent with a transferrin-like mechanism¹², whereas the uptake of Fe-ferrioxamine B was unaffected (Fig. 3a and Extended Data Fig. 5a). In Δ ISIP2A, ⁵⁹Fe³⁺ uptake remained correlated with [CO₃²⁻]; this probably reflects the activity of ISIP2A homologues (Extended Data Fig. 1b).

Iron binds to transferrin through a synergistic interaction between Fe³⁺ and CO₃²⁻, and neither substrate binds in the absence of the other¹³. This synergistic binding is a defining characteristic of transferrin¹², and in the above experiments the interaction between Fe³⁺ and CO₃²⁻ is significant ($P = 2 \times 10^{-6}$), demonstrating a second-order rate dependency (Fig. 3b). Using these as representative measurements, we derived approximations for the pseudo-first-order rate constants for the dependency of Fe uptake on Fe³⁺ (0.78 ± 0.05 mol Fe per cell per h per M Fe³⁺ (hereafter mol Fe cell⁻¹ h⁻¹ (M Fe³⁺)⁻¹) and CO₃²⁻ (0.0016 ± 0.0009 mol Fe per cell per h per M CO₃²⁻ (hereafter mol Fe cell⁻¹ h⁻¹ (M CO₃²⁻)⁻¹; Extended Data Fig. 6). Although organic

protein control. **c**, Site-directed mutagenesis of putative iron-coordinating residues. **d**, In *P. tricornutum* acclimatized to low-iron concentrations, RFP-labelled ISIP2A (red) localizes to the outer membrane and internal vesicles. Iron-regulated ISIP1 (blue) and plastid (green) are co-visualized. Scale bar, 5 μ m. **b**, **c**, Data are mean \pm s.e.m.; $n = 3$ biological replicates; * $P < 0.05$; ** $P < 0.01$; NS, not significant; paired t -test. **d**, Representative image of $n = 5$ biological replicates.

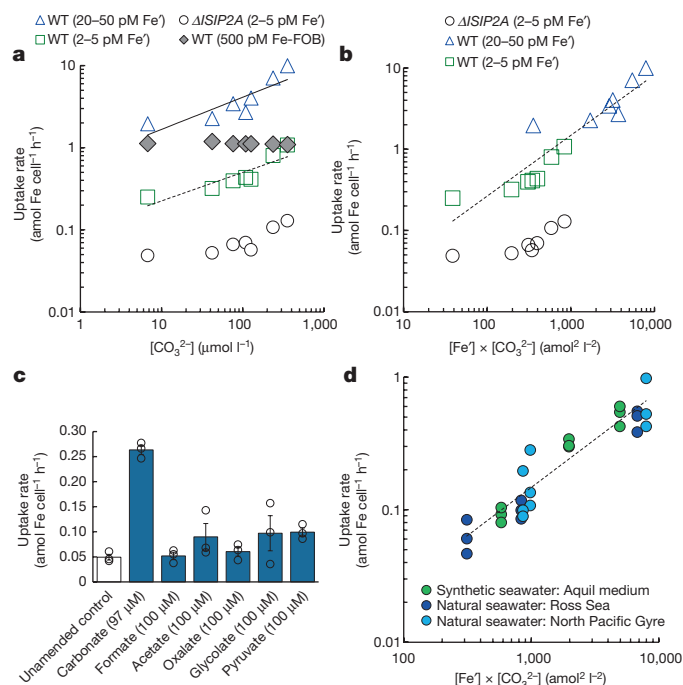


Figure 3 | The synergistic interaction of Fe' and CO₃²⁻ determines Fe' uptake rates. **a**, ⁵⁹Fe' uptake rates of *P. tricornutum* resuspended in pH-controlled synthetic seawater are correlated to [CO₃²⁻], whereas the uptake of ⁵⁹Fe-ferrioxamine B is not positively correlated. $P = 8 \times 10^{-5}$ (solid line), $P = 3 \times 10^{-5}$ (dashed) on $n = 7$ individual experiments per regression. WT, wild type. **b**, Data from **a** replotted as a function of [Fe'] \times [CO₃²⁻], revealing a synergistic interaction. Linear regression $P = 2 \times 10^{-10}$, $n = 14$ individual experiments. **c**, ⁵⁹Fe' uptake rates are not restored with organic acids, suggesting an *in vivo* specificity for carbonate. Data are mean \pm s.e.m.; $n = 3$ biological replicates; $**P < 0.01$, all others $P > 0.05$; paired *t*-test. **d**, The [Fe'] \times [CO₃²⁻] synergistic interaction is also observed in CO₂-manipulated natural seawater. Data are mean \pm s.e.m.; $n = 3$ biological replicates; $n = 27$ individual experiments; linear regression $P = 6 \times 10^{-9}$.

acids can replace CO₃²⁻ in an anion–Fe–transferrin ternary complex²², evidence for *in vivo* iron assimilation is rare²³, and a panel of surrogate anions failed to restore *P. tricornutum* uptake rates (Fig. 3c). When cells are resuspended in natural seawater, the synergistic interaction between Fe' and CO₃²⁻ is also observed, further supporting carbonate ion as the most likely Fe-coordinating anion (Fig. 3d). The correlation between Fe' uptake rates and [CO₃²⁻] occurs at environmentally relevant concentrations of CO₃²⁻, indicating that [CO₃²⁻] co-limits Fe' uptake, and suggesting that the projected decline in seawater [CO₃²⁻] due to ocean acidification could negatively affect phytotransferrin-mediated iron acquisition²⁴.

To investigate the effects of acidification on Fe' uptake rates, we aerated cultures of *P. tricornutum* with 50–5,000 p.p.m. CO₂, generating cultures in which the pH ranged from 8.7 to 7.2 (Extended Data Table 3). In seawater, acidification increases [H⁺] and decreases [CO₃²⁻], two linked processes that separately influence each component of the synergistic interaction between Fe' and CO₃²⁻. In seawater in which [Fe'] is buffered with EDTA, the increase in [H⁺] decreases the favourability of Fe–EDTA dissociation²⁵. Therefore, as pH decreases, [Fe'] decreases, and this change in iron chemistry has been documented to decrease *P. tricornutum* uptake rates²⁶. However, here we show that the interactive effect between CO₃²⁻ and Fe' in determining uptake rates is significant ($P = 0.00375$), with the [Fe'] \times [CO₃²⁻] product resulting in a better uptake model than [Fe'] alone (Fig. 4, solid line, and Extended Data Fig. 5b, c). Our results show that the effects of acidification are two-edged: whereas low pH decreases [Fe'], low [CO₃²⁻] decreases

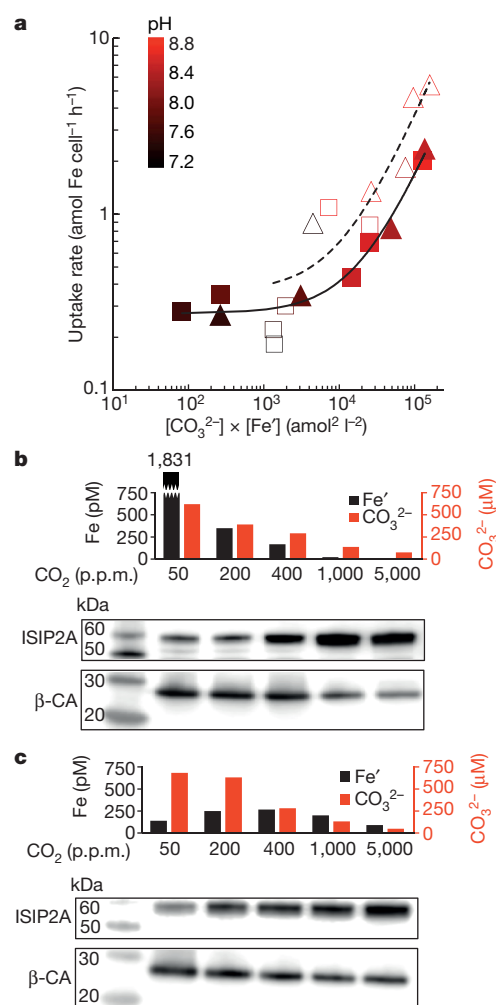


Figure 4 | pH and CO₂ manipulation of *P. tricornutum* induces expression of ISIP2A. **a**, In *P. tricornutum* cultures, elevated CO₂ lowers pH, decreasing both [CO₃²⁻] and [Fe'] and resulting in decreased ⁵⁹Fe' uptake rates (filled shapes and solid line). $n = 9$ individual experiments; linear regression $P = 5 \times 10^{-8}$. Constraining the change in [Fe'] relative to [CO₃²⁻] confirms the influence of [CO₃²⁻] on uptake rates (open shapes, dashed line). $n = 10$ individual experiments, linear regression $P = 2 \times 10^{-5}$. **b**, Immunoblots of ISIP2A and β -carbonic anhydrase (β -CA) for conditions in which the change in [CO₃²⁻] and [Fe'] is unconstrained (solid triangles from **a**). **c**, Immunoblots of ISIP2A and β -carbonic anhydrase for conditions in which the change in [Fe'] is constrained in relation to the change in [CO₃²⁻] (open triangles from **a**). Graphs indicate [Fe'], [CO₃²⁻] and bubbled CO₂ per immunoblot, β -carbonic anhydrase was used as carbon chemistry indicator. Immunoblot experiments were replicated three times with independent cultures, each with equivalent results.

Fe' bioavailability, and both of these processes act individually and in tandem to reduce phytotransferrin-mediated iron uptake rates.

High expression levels of ISIP2A are an indication of iron-limited growth⁶, and in the high-CO₂ cultures, ISIP2A expression may be responding to [Fe'], [CO₃²⁻] or both (Fig. 4b). To demonstrate that changes in [CO₃²⁻] can induce the expression of ISIP2A, we constrained the change in [Fe'] by altering the Fe:EDTA ratio (Extended Data Table 4). We show that ⁵⁹Fe uptake rates are correlated with changes in [CO₃²⁻], but not with changes in [Fe'] (Extended Data Fig. 5d, e). As in the NaHCO₃-manipulation experiments, the interaction between Fe' and CO₃²⁻ remained a significant model for predicting Fe' uptake rates (Fig. 4a, dashed line), confirming that elevated CO₂ decreases the bioavailability of Fe' and induces the expression of ISIP2A (Fig. 4c). This link between CO₂ and iron acquisition resolves the puzzling

Fe–CO₂ co-regulation of FEA1 (also known as H43), an algal iron-binding protein²⁷ that was initially identified during high-CO₂ cultivation of the chlorophyte *Chlorococcum littorale*²⁸. Our placement of FEA1 in the phytotransferrin family provides a biochemical basis for the observed influence of CO₂ on iron homeostasis while validating the CO₂ sensitivity of phytotransferrin in distantly related taxa (Fig. 1).

The requirement for synergistic binding reveals the existence of a previously undescribed form of iron–carbonate co-limitation²⁹ that may be relevant in environments in which primary productivity is limited by [Fe]. Like transferrin, phytotransferrin exploits carbonate chemistry to function as both coordination anion and internalization release trigger¹³, resulting in a mechanism that is exquisitely sensitive to acidification-induced changes in [CO₃²⁻]. Our results show that under constant [Fe'], the doubling of CO₂ to 800 p.p.m. CO₂ can reduce *P. tricornutum* Fe' uptake rates by 44%. While Fe' is an important component of phytoplankton nutrition⁵, we also show that *P. tricornutum* can use phytotransferrin-independent pathways to access organically complexed iron. In the marine environment, this pool of complexed iron is much larger³, although uptake rates can be orders of magnitude slower³⁰, underscoring the trade-off inherent to the different iron acquisition strategies. As iron-limited regions exert an important influence on global biogeochemical cycles¹, we view these results as a starting point for understanding the complex and interdependent influences of ocean acidification on phytoplankton iron uptake mechanisms and rates.

Online Content Methods, along with any additional Extended Data display items and Source Data, are available in the online version of the paper; references unique to these sections appear only in the online paper.

Received 30 January 2017; accepted 5 February 2018.

Published online 14 March 2018.

- Moore, C. M. *et al.* Processes and patterns of oceanic nutrient limitation. *Nat. Geosci.* **6**, 701–710 (2013).
- Boyd, P. W. *et al.* Mesoscale iron enrichment experiments 1993–2005: synthesis and future directions. *Science* **315**, 612–617 (2007).
- Gledhill, M. & Buck, K. N. The organic complexation of iron in the marine environment: a review. *Front. Microbiol.* **3**, 69 (2012).
- Barbeau, K., Rue, E. L., Bruland, K. W. & Butler, A. Photochemical cycling of iron in the surface ocean mediated by microbial iron(III)-binding ligands. *Nature* **413**, 409–413 (2001).
- Morel, F. M. M., Kustka, A. B. & Shaked, Y. The role of unchelated Fe in the iron nutrition of phytoplankton. *Limnol. Oceanogr.* **53**, 400–404 (2008).
- Allen, A. E. *et al.* Whole-cell response of the pennate diatom *Phaeodactylum tricornutum* to iron starvation. *Proc. Natl Acad. Sci. USA* **105**, 10438–10443 (2008).
- Lommer, M. *et al.* Genome and low-iron response of an oceanic diatom adapted to chronic iron limitation. *Genome Biol.* **13**, R66 (2012).
- Morrissey, J. *et al.* A novel protein, ubiquitous in marine phytoplankton, concentrates iron at the cell surface and facilitates uptake. *Curr. Biol.* **25**, 364–371 (2015).
- Marchetti, A. *et al.* Comparative metatranscriptomics identifies molecular bases for the physiological responses of phytoplankton to varying iron availability. *Proc. Natl Acad. Sci. USA* **109**, E317–E325 (2012).
- Bertrand, E. M. *et al.* Phytoplankton–bacterial interactions mediate micronutrient colimitation at the coastal Antarctic sea ice edge. *Proc. Natl Acad. Sci. USA* **112**, 9938–9943 (2015).
- Lambert, L. A., Perri, H., Halbrooks, P. J. & Mason, A. B. Evolution of the transferrin family: conservation of residues associated with iron and anion binding. *Comp. Biochem. Physiol.* **142**, 129–141 (2005).
- Aisen, P., Leibman, A. & Zweier, J. Stoichiometric and site characteristics of the binding of iron to human transferrin. *J. Biol. Chem.* **253**, 1930–1937 (1978).
- Baker, H. M., Anderson, B. F. & Baker, E. N. Dealing with iron: common structural principles in proteins that transport iron and heme. *Proc. Natl Acad. Sci. USA* **100**, 3579–3583 (2003).
- Fisher, M., Gokhman, I., Pick, U. & Zamir, A. A structurally novel transferrin-like protein accumulates in the plasma membrane of the unicellular green alga *Dunaliella salina* grown in high salinities. *J. Biol. Chem.* **272**, 1565–1570 (1997).
- Anderson, M. A. & Morel, F. M. The influence of aqueous iron chemistry on the uptake of iron by the coastal diatom *Thalassiosira weissflogii*. *Limnol. Oceanogr.* **27**, 789–813 (1982).
- Bruns, C. M. *et al.* Structure of *Haemophilus influenzae* Fe³⁺-binding protein reveals convergent evolution within a superfamily. *Nat. Struct. Biol.* **4**, 919–924 (1997).
- Och, L. M. & Shields-Zhou, G. A. The Neoproterozoic oxygenation event: environmental perturbations and biogeochemical cycling. *Earth Sci. Rev.* **110**, 26–57 (2012).
- Smith, S. R. *et al.* Transcriptional orchestration of the global cellular response of a model pennate diatom to diel light cycling under iron limitation. *PLoS Genet.* **12**, e1006490 (2016).
- Karas, B. J. *et al.* Designer diatom episomes delivered by bacterial conjugation. *Nat. Commun.* **6**, 6925 (2015).
- Harding, C., Heuser, J. & Stahl, P. Receptor-mediated endocytosis of transferrin and recycling of the transferrin receptor in rat reticulocytes. *J. Cell Biol.* **97**, 329–339 (1983).
- Fisher, M., Zamir, A. & Pick, U. Iron uptake by the halotolerant alga *Dunaliella* is mediated by a plasma membrane transferrin. *J. Biol. Chem.* **273**, 17553–17558 (1998).
- Schlabach, M. R. & Bates, G. W. The synergistic binding of anions and Fe³⁺ by transferrin. Implications for the interlocking sites hypothesis. *J. Biol. Chem.* **250**, 2182–2188 (1975).
- Halbrooks, P. J., Mason, A. B., Adams, T. E., Briggs, S. K. & Everse, S. J. The oxalate effect on release of iron from human serum transferrin explained. *J. Mol. Biol.* **339**, 217–226 (2004).
- Doney, S. C., Fabry, V. J., Feely, R. A. & Kleypas, J. A. Ocean acidification: the other CO₂ problem. *Annu. Rev. Mar. Sci.* **1**, 169–192 (2009).
- Sunda, W. & Huntsman, S. Effect of pH, light, and temperature on Fe–EDTA chelation and Fe hydrolysis in seawater. *Mar. Chem.* **84**, 35–47 (2003).
- Shi, D., Xu, Y., Hopkinson, B. M. & Morel, F. M. M. Effect of ocean acidification on iron availability to marine phytoplankton. *Science* **327**, 676–679 (2010).
- Allen, M. D., del Campo, J. A., Kropat, J. & Merchant, S. S. FEA1, FEA2, and FRE1, encoding two homologous secreted proteins and a candidate ferrireductase, are expressed coordinately with FOX1 and FTR1 in iron-deficient *Chlamydomonas reinhardtii*. *Eukaryot. Cell* **6**, 1841–1852 (2007).
- Sasaki, T., Kurano, N. & Miyachi, S. Cloning and characterization of high-CO₂-specific cDNAs from a marine microalga, *Chlorococcum littorale*, and effect of CO₂ concentration and iron deficiency on the gene expression. *Plant Cell Physiol.* **39**, 131–138 (1998).
- Saito, M. A., Goepfert, T. J. & Ritt, J. T. Some thoughts on the concept of colimitation: three definitions and the importance of bioavailability. *Limnol. Oceanogr.* **53**, 276–290 (2008).
- Lis, H., Shaked, Y., Kranzler, C., Keren, N. & Morel, F. M. M. Iron bioavailability to phytoplankton: an empirical approach. *ISME J.* **9**, 1003–1013 (2015).

Supplementary Information is available in the online version of the paper.

Acknowledgements We thank J. Badger for early contributions to phylogenetic analyses, A. Dickson for pH analysis, K. Forsch for CSV measurements and E. Bertrand for trace-metal clean techniques. This study was supported by the National Science Foundation (NSF-MCB-1024913, NSF-ANT-1043671 and NSF-OCE-0727997), United States Department of Energy Genomics Science program (DE-SC00006719 and DE-SC0008593), and the Gordon and Betty Moore Foundation grant GBMF3828 (A.E.A.); NSF-1557928 (A.B.K.); and the Czech Science Foundation, project 15-17643S (M.O. and A.H.).

Author Contributions J.B.M., A.B.K., M.O. and A.E.A. designed the study and interpreted the results. J.B.M., M.O. and A.H. generated and analysed phylogenetic and molecular clock data. J.B.M. and B.J.K. generated mutant cell lines, J.B.M. and A.B.K. with assistance from K.A.B. performed physiology experiments. J.B.M. performed microscopy, and H.Z. performed western blots. T.K. and A.J.A. analysed inorganic carbon species. J.B.M. and J.P.M. conducted statistical analyses. J.B.M. wrote the paper with input from A.E.A., A.B.K., M.O., J.P.M., A.J.A. and K.A.B. All authors discussed the results and commented on the manuscript.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing interests. Readers are welcome to comment on the online version of the paper. Publisher's note: Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations. Correspondence and requests for materials should be addressed to A.E.A. (aallen@jcvl.org).

Reviewer Information Nature thanks S. Amin, E. DeLong and the other anonymous reviewer(s) for their contribution to the peer review of this work.

METHODS

Phylogenetic analysis. To maximize the dimensions of our phylogenetic analysis, we built a series of hidden Markov models (HMM) from manually curated multiple peptide alignments. HMMs were assembled using transferrin and ISIP2A sequences obtained from NCBI via BLAST. Transferrin and ISIP2A HMMs were then used to iteratively obtain sequences from the Marine Microbial Eukaryote Transcriptome Sequencing Project (MMETSP)³¹. For inclusion into the transferrin and phytotransferrin clades, sequences were manually screened to ensure that they contained the conserved metal-binding residues and the carbonate-binding region (Extended Data Fig. 1a). Sequences were aligned using the genafpair exhaustive algorithm included in MAFFT³². The alignment was then manually inspected and ambiguous regions were removed in SEAVIEW v.4³³. Phylogenetic analysis was conducted using maximum likelihood under the gamma-corrected LG matrix in RAXML v.8.2.8³⁴. The best topology, as well as non-parametric branching support was inferred using fast algorithm (-fa option) from 1,000 replicates. Owing to the high divergence of the sequences, we have alternatively used Bayesian inference under the probabilistic CAT algorithm with numbers of site categories limited to 40 (C40 model), combined with empirical exchange rates as defined by the LG model in PhyloBayes v.4.1³⁵. Two independent Markov chain Monte Carlo (MCMC) chains were run until they converged (that is, maximum observed discrepancy was below 0.2) and minimum effective size of model parameters exceeded 100.

Divergence time estimation. Fossil calibration points for multicellular taxa (metazoans and green plants) were compiled from <http://fossilcalibrations.org> (Extended Data Table 4). Calibration of diatoms and haptophytes (mineralizing protozoans) was adopted according previously published studies^{36,37}. Respective values are summarized in Supplementary Table 4. The best-scoring topology from the Bayesian inference (see above) was then used for an estimation of divergence using UGAM and CIR relaxed clock models in PhyloBayes v.4.1. In addition, we used an uncorrelated log-normal clock model as implemented in BEAST v.2³⁸. For this model, MCMC was run for 30 million generations. Convergence as well as effective sample size was evaluated using Tracer and the dating visualized on the tree using TreeAnnotator in the BEAST package.

Predictions of signal peptides, transmembrane domains and glycosylphosphatidylinositol anchors. All the sequences used for phylogenetic reconstructions were searched for the presence of signal peptide, transmembrane domains and glycosylphosphatidylinositol anchor (GPI). Endoplasmic reticulum signal peptides were predicted using SignalP³⁹ and TargetP^{40,41}. Transmembrane domains were searched using TMHMM⁴² and GPI anchors were predicted by PredGPI⁴³. The results are shown in Supplementary Table 2.

Cell lines and cultivation. All cultivation and manipulation of *P. tricornutum* strain CCMP 632 was conducted using sterile trace-metal clean techniques and in Aquil medium⁴⁴, unless otherwise noted. To remove contaminating iron, prepared medium was first passed through a Chelex 100 resin (Bio-Rad Laboratories) and microwave-sterilized in acid-cleaned polycarbonate bottles before addition of filter-sterilized, Chelex-passaged nutrient stocks and F/2 vitamin supplements.

Cultures for short-term ⁵⁹Fe uptake rate experiments. Cultures were grown in medium supplemented with macronutrients (880 μM NO₃⁻, 36 μM PO₄³⁻, 106 μM SiO₄²⁻) and trace metals (100 nM Zn, 48 nM Mn, 40 nM Co, 40 nM Cu, 10 nM Se and 100 nM Ni) with 100 μM EDTA. Vitamin and metal-EDTA solutions were added to artificial seawater and allowed to equilibrate overnight. Iron was pre-equilibrated overnight in a 1:1.25 ratio of iron:EDTA before addition. Cultures were maintained at 17 °C at 300 μmol quanta m⁻² s⁻¹ in a 10-h:14-h dark:light cycle, and all culture manipulations were done under a class-100 HEPA filter. Background contaminating iron was measured by flow injection analysis at multiple times and varied from 0.2–0.7 nM dissolved iron over the two-year course of experiments. Fe³⁺, the sum of all labile inorganic iron species, was calculated as previously described⁴⁴.

Cultures grown for growth rate analysis. Cultures of *P. tricornutum* (CCMP 632) were grown in a modified Aquil medium⁴⁴ containing EDTA, Co, Cu, Mo, Mn and Zn at final concentrations of 3×10^{-4} M, 1.51×10^{-7} M, 1×10^{-7} M, 3.36×10^{-7} M, 1×10^{-8} M and 2.39×10^{-7} M, with variable Fe concentrations. Cells were grown at 320 μmol photons m⁻² s⁻¹ under a 12-h:12-h light:dark cycle using cool-white fluorescent bulbs. Under these conditions, Fe³⁺ = 0.00068 of total Fe. A maximum of 700 pmol l⁻¹ Fe³⁺ was assumed, because of iron hydroxide precipitation, but for simplicity, we report the results as Fe³⁺ even at concentrations exceeding this maximum. Cell growth was monitored daily using a Turner AU-10 fluorometer for *in vivo* chlorophyll fluorescence, and steady-state growth rates were computed from linear regressions of ln(relative fluorescence) versus time.

TALEN knockouts, transferrin complementation and site-directed mutagenesis. Transcription activator-like effector nucleases (TALENs) were used to disrupt the *ISIP2A* gene. TALEN construction protocols followed the methods

as described for *P. tricornutum*⁴⁹. In brief, a unique pair of 20-nucleotide length sequences was designed to span the first intron junction of ISIP2A (GenBank ID: XP_002179762.1). TALEN constructs were assembled into the pTH vector⁴⁹ and the sequence was verified. A homologous recombination (HR) plasmid containing resistance to nourseothricin and driven by the promoter/terminator for the light harvesting promoter *FcpA*⁴⁵ was assembled into pUC-19 using Gibson cloning⁴⁶. Both HR and pTH plasmids were introduced into *P. tricornutum* via particle bombardment⁴⁷. Transformants were pre-screened by PCR for correct insertion of the nourseothricin-resistance cassette and Δ ISIP2A gene knockouts were subsequently confirmed by western blot using custom-made Genomic Antibodies (OriGene) targeting the ISIP2A protein. Complementary sequences for the N- and C-terminal domains of hTF were codon-optimized for the diatom *Thalassiosira pseudonana* and designed using CLC software (Qiagen), and the signal peptide and C-terminal transmembrane anchor of ISIP2A were subsequently added before synthesis. The fusion products were then synthesized on a BioXp 3200 DNA synthesizer (SGI DNA). YFP was fused via Gibson cloning to the signal and transmembrane sequences of ISIP2A, while the ISIP2A complement was obtained by amplifying *P. tricornutum* cDNA (GeneScript). PCR-based site-directed mutagenesis strains were made with base-pair mismatch primers and amplified from the ISIP2A complement strain. All constructs were placed under control of the *P. tricornutum* nitrate reductase promoter and terminator^{19,48}, sequence-verified, and maintained on pUC-19 vectors ($n = 3$ for each construct). All complement sequences were Gibson-assembled into the cargo plasmid p0521s¹⁹ (containing a selection marker for bleomycin) and introduced into the Δ ISIP2A-knockout strain via conjugation with *Escherichia coli* Epi300 cells as previously described⁴⁹. Positive transformants were selected on dual nourseothricin/zeocin antibiotic plates (200 and 100 μg ml⁻¹, respectively) and verified by PCR and sequencing. Cell lines were maintained in 7.5 nM Fe-EDTA Aquil medium under positive antibiotic selection until testing.

Short-term ⁵⁹Fe uptake assays. Cell strains were grown without antibiotic selection and acclimatized in Aquil medium at 15 pM Fe³⁺ for a minimum of five transfers before assay. Growth in 15 pM Fe³⁺ was found to reduce the growth rate in wild-type *P. tricornutum* by approximately 15%, with an approximate F_v/F_m of 0.55, indicative of iron-limited (but not iron-starved) growth⁶. Cell counts were conducted on a Beckman Quanta cell cytometer and cross-calibrated by direct microscopic counts using a haemocytometer. To avoid changes to physiology and medium pH, cultures were grown and assayed at low cell densities ($1\text{--}3 \times 10^5$ cells ml⁻¹). ⁵⁹Fe was equilibrated with EDTA at a 1:2 ratio and pH 7 for 24 h before use. The specific activity of the ⁵⁹Fe was 42 Ci g⁻¹, or about 0.08% pure. Desferrioxamine B (DFB) was complexed with ⁵⁹Fe by equilibrating at pH 3 overnight and adjusting to pH 8 with high-purity NaOH. The ratio of ⁵⁹Fe to uncomplexed DFB was 4:5. Precomplexed ⁵⁹Fe-EDTA and ⁵⁹Fe-DFB substrates were added to standard medium containing 100 μM EDTA and allowed to equilibrate overnight. Uptake assays were commenced by gravimetrically adding a 1:10 volume to the cell cultures. Uptake assay data were generated from uptake curves consisting of a minimum of five time points spaced 20–30 min apart. Assays were simultaneous and staggered, resulting in variations in the total length of assay; however, assays were no less than 2 h in length, and did not exceed 3 h in duration. ISIP2A protein expression has a strong diurnal pattern: for consistency and reproducibility, cells were collected in the light portion of the light:dark cycle. Room temperature fluctuations influence both carbon and iron speciation; temperature changes were accounted for in calculating Fe:EDTA dissociation constants²⁵ and CO₂ speciation⁵⁴. During uptake assays, cells were shielded from actinic light and periodically filtered onto 2.0-μm PTFE filters, washed with Ti-citrate-EDTA solution⁵⁰ and preserved in Ecolite (MP Biomedicals) before liquid scintillation counting.

Carbonic acid system manipulation and surrogate anion experiments. Medium for the carbonic acid manipulation experiments was made by acidifying Chelex-treated bicarbonate-free Aquil medium to pH 3 with ultra-high purity HCl and bubbling with filtered N₂ for 60 min to strip out dissolved CO₂. To buffer against large (± 0.1 pH) changes in pH, Chelex-treated Tris-HCl was added at a final concentration of 2 mM, and the pH was adjusted back to 8.0 using ultra-high purity ammonium hydroxide. Ammonium hydroxide was used rather than sodium hydroxide to avoid contaminating the uptake medium with sodium bicarbonate, which is a common atmospherically derived contaminant in NaOH. Trace-metal clean F/2 nutrients were added, and all medium was stored in an N₂/O₂-purged, CO₂-impermeable glove box before use. Carbonate ion availability was manipulated by adding NaHCO₃ to sub-aliquoted bottles of medium at concentrations of 0, 0.5, 1, 1.5, 2, 4 and 6 mM NaHCO₃ l⁻¹, and allowed to equilibrate 24 h. ⁵⁹Fe-EDTA or ⁵⁹Fe-FOB were added 24 h before assay and allowed to equilibrate. *P. tricornutum* cells were grown in 2.7-l square polycarbonate bottles in Aquil medium containing 15 pM Fe³⁺ (7.5 nM Fe-EDTA), and were grown on a 10-h:14-h light:dark schedule. Cultures were grown to early exponential phase (3×10^5 cells ml⁻¹) and collected

by centrifugation for 30 min at 1,000g at 17°C in acid-cleaned centrifuge bottles. Cells were washed and centrifuged twice in NaHCO_3 -free Aquil synthetic seawater salts (without nutrients), and re-suspended in the carbonate chemistry-adjusted ^{59}Fe medium. Uptake reaction volume was 150 ml, and to avoid unnecessary aeration, the medium was removed from the medium bottles using an acid-cleaned pipette. Volumes were determined gravimetrically. Uptake rates were assayed as outlined above (5-point curves, 2–3 h in duration), and samples for pH and total dissolved inorganic carbon (DIC) analysis removed from parallel (non-radioactive) cultures, poisoned with HgCl_2 , sealed with Apiezon-L vacuum grease (M&I Materials) and stored in borosilicate-stoppered bottles for analysis⁵¹. Owing to the presence of Tris-HCl, total alkalinity was not analysed for this experiment and inorganic carbon speciation was calculated from pH and DIC (see below).

To test whether other anions could substitute for carbonate ions, bicarbonate-free medium was prepared as above. Sub-aliquots of this medium were used to create 100 mM stocks of the sodium salts of formate, acetate, oxalate, glycolate and pyruvate. These stocks were Chelex-treated for 48 h at 4°C before use. These alternate anion stocks were diluted 1:1,000 into 150-ml uptake medium reactions and the medium was allowed to equilibrate in an N_2/O_2 -purged, CO_2 -impermeable glove box as above. Cells were grown, collected and resuspended as above, and added to the 150-ml uptake reactions. Uptake rates were assayed as in the previous carbonic acid manipulation experiments: 5-point assay curves with a total uptake assay length of 2–3 h. Samples for pH and DIC analysis were removed from parallel (non-radioactive) cultures, poisoned with HgCl_2 , sealed with Apiezon-L vacuum grease (M&I Materials) and stored in borosilicate-stoppered bottles for analysis⁵¹. Carbonate contamination in the unsupplemented control was 2.5 μM (Extended Data Table 2).

Demonstration of carbonate effect in natural seawater. On 16 January 2013, seawater was collected from 3 m depth at the edge of the sea ice in McMurdo Sound Ross Sea, Antarctica (77° 36.999' S 165° 28.464' E). Water was pumped to the surface using a trace-metal clean diaphragm pump via acid-cleaned Teflon tubing and filtered on-site through an acid-cleaned 0.22- μm filter. Filtered seawater was dispensed into trace-metal clean 50-l carboys and stored in the dark at 4°C. Phytoplankton grown in this batch of seawater are iron-limited¹⁰. For North Pacific Gyre water, seawater from station ALOHA (22° 45' N 158° 00' W) was collected in June 2016 using trace-metal clean techniques and similarly filtered and cold-stored before use. Samples of both types of natural seawater plus synthetic seawater (Aquil) were dispensed into 2.7-l acid-cleaned polycarbonate bottles. Seawater was amended with 100 μM EDTA and Chelex-treated F/2 nutrients and allowed to equilibrate overnight. Two-litre samples of seawater were placed in 2.7-l acid-cleaned polycarbonate bottles and bubbled with filtered N_2/O_2 mixtures containing 200 and 1,000 p.p.m. CO_2 . Tubing and lines were acid-cleaned PTFE, with acid-soaked and -cleaned plastic aerators (Lee's Discard-a-Stone) used to maximize gas transfer. One hundred and fifty millilitres of each seawater sample (CO_2 -modified samples and the unmodified controls) was transferred to small polycarbonate bottles as uptake medium, with 7.5 nM ^{59}Fe -EDTA allowed to equilibrate overnight. Separately, *P. tricornutum* was grown in 7.5 nM Fe-EDTA Aquil to early exponential phase (3×10^5 cells ml^{-1}) and harvested by centrifugation for 30 min at 1,000g at 17°C in acid-cleaned centrifuge bottles. This concentrate was resuspended in the uptake medium, and uptake rates were assayed as outlined above (5 point curves, 2–3 h in duration), and samples for total alkalinity and DIC analysis removed from parallel (non-radioactive) cultures, poisoned with HgCl_2 , sealed with Apiezon-L vacuum grease (M&I Materials), and stored in borosilicate-stoppered bottles for analysis⁵¹. This experiment was repeated with three cultures on three separate days.

Approximation of Fe' and carbonate rate constants (NaHCO_3 addition experiments). The short-term uptake rates determined from an iron-limited culture incubated across a range of $[\text{CO}_3^{2-}]$ and Fe' levels (Extended Data Table 1) were used to determine a second-order rate constant for Fe' uptake and to solve for pseudo-first-order rate constants with respect to Fe' and $[\text{CO}_3^{2-}]$. These results are expressed on a per-cell basis, and pertain to cultures of *P. tricornutum* grown under this condition⁵¹, as the starting ISIP2A protein content per cell was identical (but unknown) in all short-term uptake experiments. Uptake rates determined from the two treatments to which no NaHCO_3 was added were excluded. The slope of uptake versus the product of Fe' and CO_3^{2-} (1.22×10^{-3} mol Fe cell $^{-1}$ h $^{-1}$ (M Fe') $^{-1}$ (M CO_3^{2-}) $^{-1}$; Extended Data Fig. 6a), was used as the starting condition from which an initial CO_3^{2-} rate constant was determined. Assuming Fe uptake rate is first order with respect to Fe' at low ^{59}Fe (and supported by the ratios of rates observed at varied Fe' at identical $[\text{CO}_3^{2-}]$), the uptake rates measured at Fe' ranging from 2 to 45 pM (Extended Data Fig. 6b) and at variable $[\text{CO}_3^{2-}]$ were collapsed by dividing these rates by the Fe' levels (Extended Data Fig. 6d). The slope of this quantity versus $[\text{CO}_3^{2-}]$ revealed an initial constant for Fe' -normalized Fe uptake of 1.27×10^{-3} mol Fe cell $^{-1}$ h $^{-1}$ (M CO_3^{2-}) $^{-1}$. Then, a pseudo-first-order rate

constant of Fe' for Fe uptake, $k = 0.962$ mol Fe cell $^{-1}$ h $^{-1}$ (M Fe') $^{-1}$, was estimated from the second-order rate constant and the estimated CO_3^{2-} rate constant. The constituent rate constants for Fe uptake were solved by iteration using these initial conditions. Each round was performed with 16,000 pairs of trial rate constants, allowed to vary randomly from 0.33 to threefold, from the initial or previously iterated values. Residual sum of square values were calculated for each pair of trial rate constants. The median constants from the 20 pairs of trial constants with the lowest residual sum of square values (which had coefficients of variation of 0.01% or less) were used for the next trial. Trials were stopped when constants did not appreciably differ and no trend with trial number was apparent. For each constant, the mean of the last four trials are reported; the iterated constants were 0.782 mol (± 0.05) mol Fe cell $^{-1}$ h $^{-1}$ (M Fe') $^{-1}$ and 0.00156 (± 0.00087) mol Fe cell $^{-1}$ h $^{-1}$ (M CO_3^{2-}) $^{-1}$ for Fe' and CO_3^{2-} , respectively (about 20% different from initial estimates). The fit of the data using the derived rate constants versus empirically measured rates is shown in Extended Data Fig. 6e.

Carbonic acid system manipulation by CO_2 bubbling, synthetic seawater. Two litres of Chelex-treated Aquil medium in 2.7-l polycarbonate bottles was bubbled with filtered N_2/O_2 mixtures containing 50, 200, 1,000 and 5000 p.p.m. CO_2 plus an ambient air (~ 400 p.p.m.) control. Tubing and lines were acid-cleaned PTFE, with acid-soaked and -cleaned plastic aerators (Lee's Discard-a-Stone) used to maximize gas transfer. In the first experiment (Fe' unconstrained), both high (100 nM Fe-EDTA) and low (10 nM Fe-EDTA) Fe' media were pre-equilibrated with bubbled gas for 48 h before inoculation with a low-iron acclimatized culture of wild-type *P. tricornutum* cells. Cells were monitored twice daily for cell counts, chlorophyll and F_v/F_m . Cultures were grown to a low cell density ($< 3 \times 10^5$ cells ml^{-1}) before the uptake assay. The specific growth rates for each culture were determined from the last three cell counts. ^{59}Fe -EDTA was pre-equilibrated overnight in filter-sterilized medium from each culture to assure that medium chemistry would be similar. Assays were a mix of 135 ml culture and 15 ml ^{59}Fe uptake media, blended without aeration, periodically filtered onto 2.0- μm PTFE filters, washed with Ti-citrate-EDTA solution⁵⁰ and preserved in Ecolite (MP Biomedicals) before liquid scintillation counting. Short-term uptake assays consisted of a 5-point curve, sampled 20 min apart, and not exceeding 3 h in duration. Culture aliquots of 500 ml were preserved for DIC and total alkalinity, and the rest of the culture was centrifuged and stored for protein expression analysis. Both ISIP2A and β -carbonic anhydrase were analysed on protein-per-cell normalized samples (Invitrogen).

For the Fe' -constrained experiments, the Fe:EDTA ratio was varied to constrain pH-induced changes in $[\text{Fe}']$. For the low-iron cultures, we attempted to constrain Fe to ~ 20 pM Fe' across all five pH and CO_2 manipulations, and in the high-iron cultures, Fe' was constrained to ~ 200 pM Fe' . To keep the maximum concentration of EDTA to 1 mM or less, the total concentration of Fe-EDTA was decreased to 50 nM in the Fe-replete cultures, and 10 nM in the Fe-limited cultures. Cultures were analysed as above.

Characterization of endocytosis. To generate fluorescently labelled cell lines, full-length ISIP2A and ISIP1 genes (including the 600 base-pair native promoter and 300-bp native terminator regions) were amplified with Phusion high-fidelity polymerase (New England Biolabs). RFP and YFP protein tags were fused onto constructs using Gibson cloning and assembled into a pUC-19 vector. Full-length constructs were re-amplified, sequence-verified and assembled into p0521s using yeast assembly²². Clones were PCR-screened, with positive clones re-sequenced. For microscopy, cells were grown at a low density (2×10^5 cells ml^{-1}) in 7.5 nM Fe-EDTA Aquil medium, centrifuged briefly and resuspended in Aquil medium containing 300 nM Fe-EDTA. Cells were immediately visualized on a Leica SP5 confocal microscope equipped with a 100 \times oil immersion objective. A 514-nm laser was used to visualize YFP, RFP and plastid autofluorescence, with emission monitored at 520–550 nm, 610–635 nm and 700–740 nm, respectively. The membrane-impermeable dye FM 1-43 was used to stain the outer membrane; wild-type *P. tricornutum* was grown at 7.5 nM Fe-EDTA, centrifuged briefly and resuspended in Aquil medium containing 300 nM Fe-EDTA and 5 $\mu\text{g ml}^{-1}$ of FM 1-43 (Thermo Fisher). Cells were visualized within 10 min of dye addition (excitation 488 nm, emission 610 nm) by confocal microscopy. To evaluate the effect of endocytosis inhibitors, we added 5 μM of the clathrin-mediated endocytosis inhibitor Pitstop2⁵³ (N-[5-(4-bromobenzylidene)-4-oxo-4,5-dihydro-1,3-thiazol-2-yl]-naphthalene-1-sulfonamide, Abcam) to short-term uptake medium containing 7.5 nM Fe-EDTA, mixed and assayed immediately, as outlined above. The toxic effects of Pitstop2 on *P. tricornutum* over the course of the 2-h uptake experiment were not evaluated.

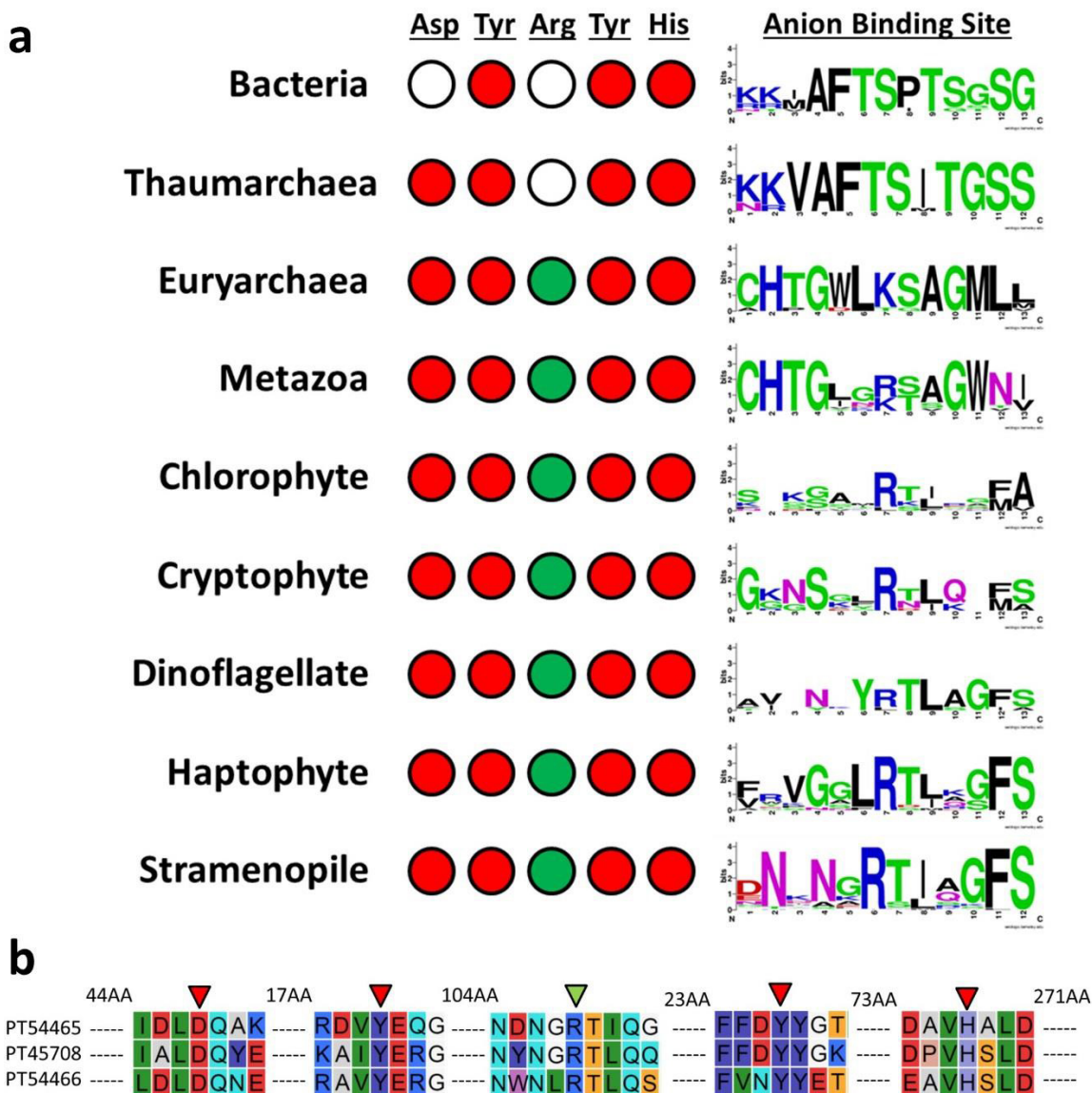
Analysis of seawater carbonate chemistry. Medium samples for pH, total alkalinity and DIC were collected following previously described protocols⁵¹. All samples were collected in 500-ml Pyrex borosilicate glass bottles and poisoned with 200 μl of a saturated HgCl_2 solution to arrest metabolic activity, and bottles were sealed using Apiezon-L sealant. For the NaHCO_3 -manipulation experiments, pH and

DIC were measured: pH was measured on the total pH scale using meta-cresol purple⁵⁴, and DIC analyses were carried out using an automated infrared inorganic carbon analyser (AIRICA, Marianda). For the CO₂ bubbling experiments, DIC and total alkalinity were measured. Total alkalinity was determined using an open-cell potentiometric acid titration system equipped with a Metrohm 876 Dosimat Plus automated titrator and a Metrohm Ecotrode Plus pH electrode, as developed by the Dickson Laboratory at Scripps Institution of Oceanography. Performance and accuracy of the analyses was evaluated through analysis of Certified Reference Material provided by the Dickson laboratory. Accuracy and precision $\pm 2 \mu\text{mol kg}^{-1}$ was achieved for both DIC and total alkalinity analyses. The complete medium carbonic acid system including carbonate ion concentration was calculated based on pH and DIC (NaHCO₃ addition experiment) or DIC and total alkalinity (CO₂ bubbling experiment) data based on *in situ* temperature and salinity using CO2SYS⁵⁵.

Statistical analyses. No statistical methods were used to predetermine sample size. All *n* numbers represent biological replicates, and the experiments were not randomized or blinded. All the data in bar graphs are expressed as mean \pm s.e.m., with paired Student's *t*-test (two-tailed). The ANOVA interaction effect between variables was used for evaluating synergism⁵⁶. Statistical analysis was performed in R⁵⁷, using linear regression (lm), two-way ANOVA (aov) and correlation tests (cor.test), where appropriate, to compare the relationship of Fe' uptake to [Fe'], [CO₃²⁻] and the [Fe'] \times [CO₃²⁻] product. Test statistics and *P* values are summarized in Extended Data Table 3. *P* < 0.05 was considered significant.

Data availability. The authors declare that the data supporting the findings of this study are available within the paper and its Supplementary Information.

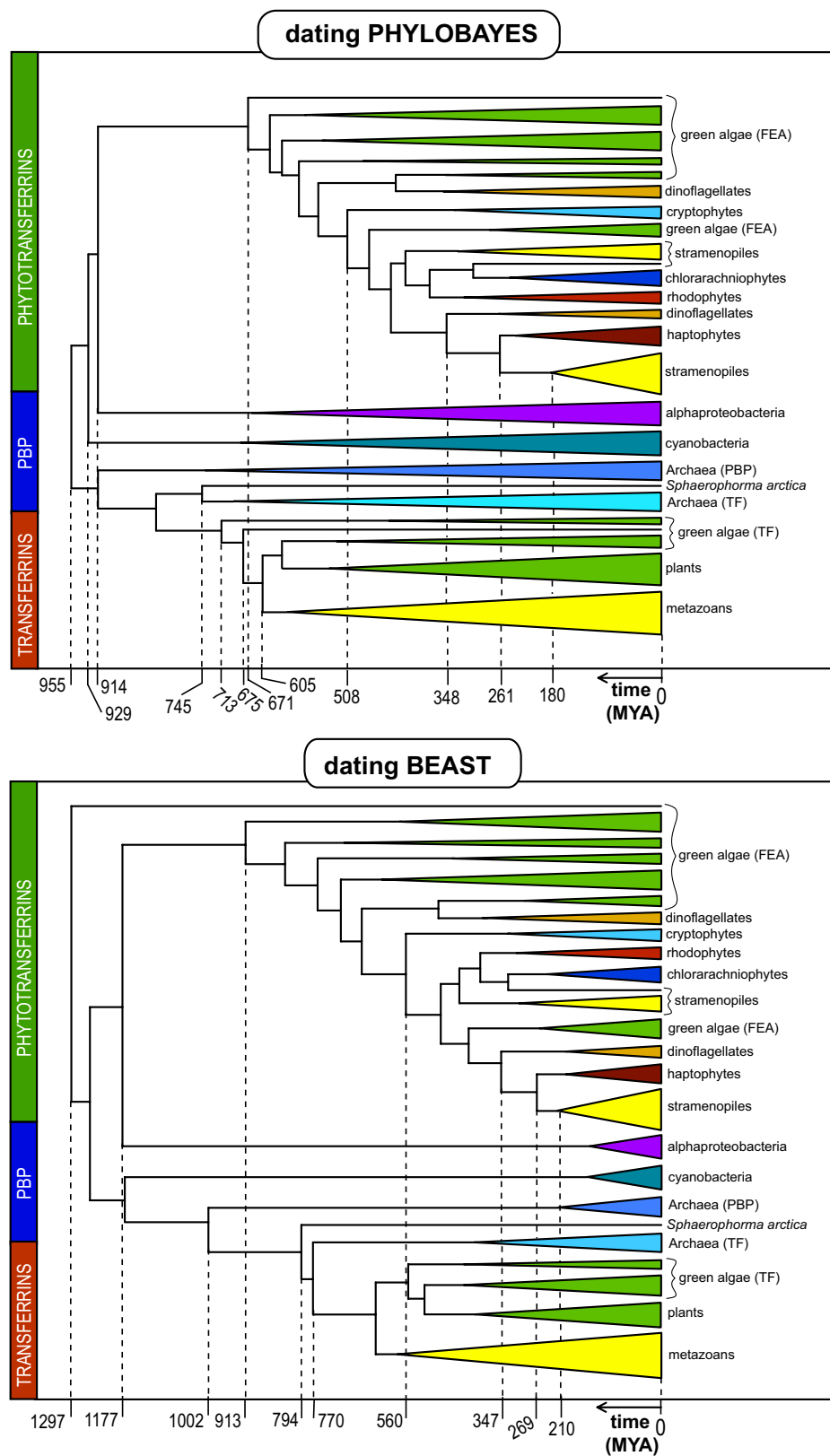
31. Caron, D. A. *et al.* Probing the evolution, ecology and physiology of marine protists using transcriptomics. *Nat. Rev. Microbiol.* **15**, 6–20 (2017).
32. Katoh, K. & Standley, D. M. MAFFT multiple sequence alignment software version 7: improvements in performance and usability. *Mol. Biol. Evol.* **30**, 772–780 (2013).
33. Gouy, M., Guindon, S. & Gascuel, O. SeaView version 4: a multiplatform graphical user interface for sequence alignment and phylogenetic tree building. *Mol. Biol. Evol.* **27**, 221–224 (2010).
34. Stamatakis, A. RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics* **30**, 1312–1313 (2014).
35. Lartillot, N., Lepage, T. & Blanquart, S. PhyloBayes 3: a Bayesian software package for phylogenetic reconstruction and molecular dating. *Bioinformatics* **25**, 2286–2288 (2009).
36. Berney, C. & Pawlowski, J. A molecular time-scale for eukaryote evolution recalibrated with the continuous microfossil record. *Proc. R. Soc. B* **273**, 1867–1872 (2006).
37. Parfrey, L. W., Lahr, D. J., Knoll, A. H. & Katz, L. A. Estimating the timing of early eukaryotic diversification with multigene molecular clocks. *Proc. Natl Acad. Sci. USA* **108**, 13624–13629 (2011).
38. Drummond, A. J., Suchard, M. A., Xie, D. & Rambaut, A. Bayesian phylogenetics with BEAUti and the BEAST 1.7. *Mol. Biol. Evol.* **29**, 1969–1973 (2012).
39. Petersen, T. N., Brunak, S., von Heijne, G. & Nielsen, H. SignalP 4.0: discriminating signal peptides from transmembrane regions. *Nat. Methods* **8**, 785–786 (2011).
40. Emanuelsson, O., Nielsen, H., Brunak, S. & von Heijne, G. Predicting subcellular localization of proteins based on their N-terminal amino acid sequence. *J. Mol. Biol.* **300**, 1005–1016 (2000).
41. Nielsen, H., Engelbrecht, J., Brunak, S. & von Heijne, G. Identification of prokaryotic and eukaryotic signal peptides and prediction of their cleavage sites. *Protein Eng.* **10**, 1–6 (1997).
42. Möller, S., Croning, M. D. R. & Apweiler, R. Evaluation of methods for the prediction of membrane spanning regions. *Bioinformatics* **17**, 646–653 (2001).
43. Pierleoni, A., Martelli, P. L. & Casadio, R. PredGPI: a GPI-anchor predictor. *BMC Bioinformatics* **9**, 392 (2008).
44. Sunda, W. G., Price, N. M. & Morel, F. M. M. in *Algal Culturing Techniques* (ed. Anderson, R. A.) 35–63 (Elsevier, 2005).
45. Zaslavskaja, L. A., Lippmeier, J. C., Kroth, P. G., Grossman, A. R. & Apt, K. E. Transformation of the diatom *Phaeodactylum tricornutum* (Bacillariophyceae) with a variety of selectable marker and reporter genes. *J. Phycol.* **36**, 379–386 (2000).
46. Gibson, D. G. *et al.* Enzymatic assembly of DNA molecules up to several hundred kilobases. *Nat. Methods* **6**, 343–345 (2009).
47. Falcione, A., Casotti, R., Leblanc, C., Abrescia, C. & Bowler, C. Transformation of nonselectable reporter genes in marine diatoms. *Mar. Biotechnol.* **1**, 239–251 (1999).
48. Poulsen, N. & Kröger, N. A new molecular tool for transgenic diatoms: control of mRNA and protein biosynthesis by an inducible promoter–terminator cassette. *FEBS J.* **272**, 3413–3423 (2005).
49. Weyman, P. D. *et al.* Inactivation of *Phaeodactylum tricornutum* urease gene using transcription activator-like effector nuclease-based targeted mutagenesis. *Plant Biotechnol. J.* **13**, 460–470 (2015).
50. Hudson, R. J. M. & Morel, F. M. M. Distinguishing between extra- and intracellular iron in marine phytoplankton. *Limnol. Oceanogr.* **34**, 1113–1120 (1989).
51. Dickson, A. G., Sabine, C. L., & Christian, J. R. (eds) *Guide to Best Practices for Ocean CO₂ Measurements* (North Pacific Marine Science Organization, 2007).
52. Kustka, A. B., Allen, A. E. & Morel, F. M. Sequence analysis and transcriptional regulation of iron acquisition genes in two marine diatoms. *J. Phycol.* **43**, 715–729 (2007).
53. Dutta, D., Williamson, C. D., Cole, N. B. & Donaldson, J. G. Pitstop 2 is a potent inhibitor of clathrin-independent endocytosis. *PLoS ONE* **7**, e45799 (2012).
54. Carter, B. R., Radich, J. A., Doyle, H. L. & Dickson, A. G. An automated system for spectrophotometric seawater pH measurements. *Limnol. Oceanogr. Methods* **11**, 16–27 (2013).
55. Lewis, E., Wallace, D. & Allison, L. J. *Program developed for CO₂ System Calculations* (Carbon Dioxide Information Analysis Center, 1998).
56. Slinker, B. K. The statistics of synergism. *J. Mol. Cell. Cardiol.* **30**, 723–731 (1998).
57. R Core Team. R: A Language and Environment for Statistical Computing. <http://www.R-project.org/> (R Foundation for Statistical Computing, Vienna, Austria, 2014).



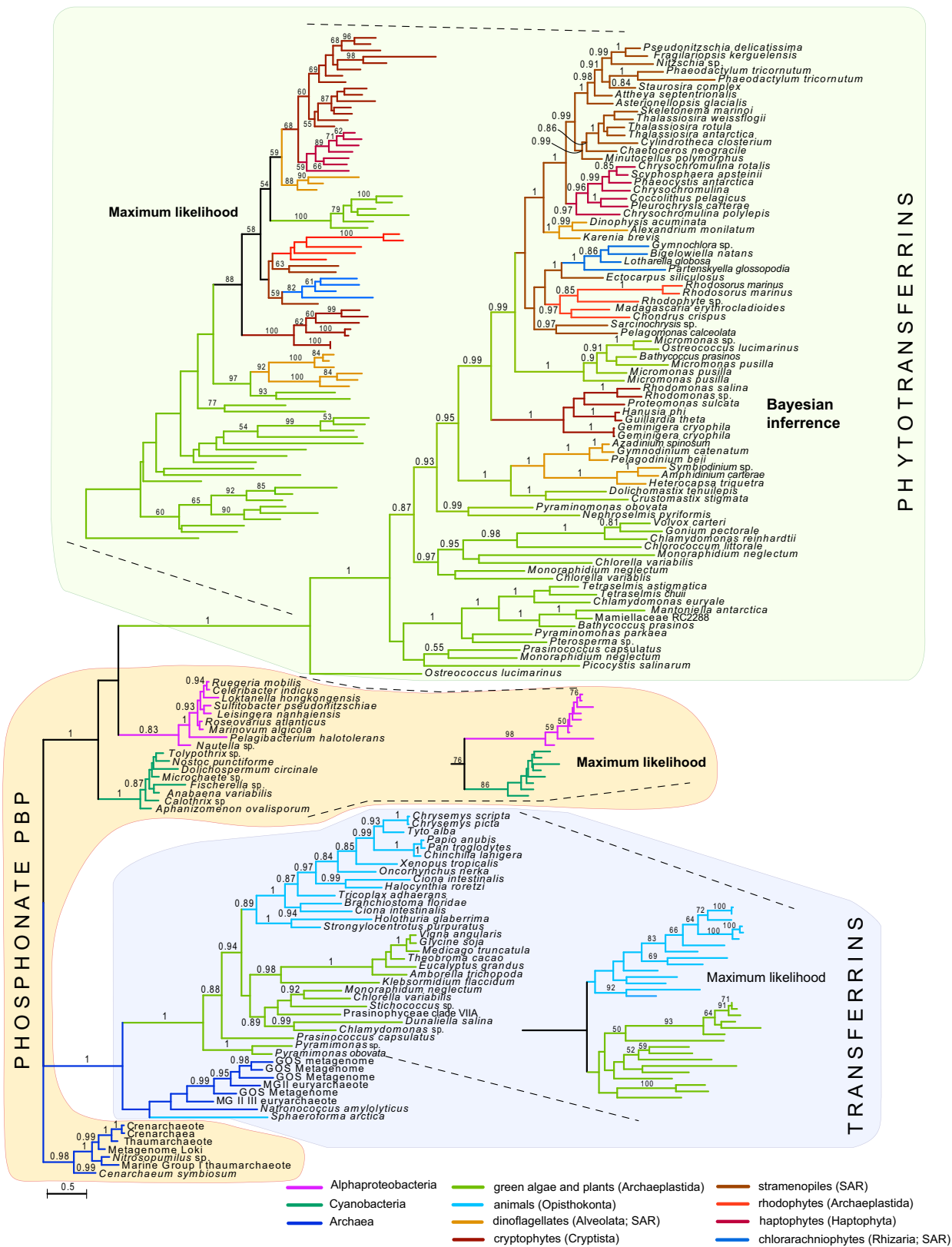
Extended Data Figure 1 | Conservation of active site amino acids.

a, Left, conservation of putative iron-coordinating (red) and carbonate-coordinating (green) amino acids among phylogenetic groups, filled circles indicate >85% conservation, unfilled circles indicate <20% conservation; arginine and lysine were permitted at the carbonate-binding site¹¹. Right, logo tag detailing alignment conservation at the anion-binding region. Phosphonate-binding proteins (Bacteria and Thaumarchaea) retain the S/T-rich phosphonate-binding area, whereas

transferrin (Euryarchaea and Metazoa) and phytotransferrin have the carbonate-coordinating K/R insertion. A downstream SAG region used to stabilize carbonate in transferrin²⁴ shows some conservation in phytotransferrin, but an upstream threonine (−4 amino acids from the conserved Arg) is not present. **b**, Active site homology among ISIP2A (PT54465) and two transmembrane-anchored *P. tricornutum* homologues, PT45708 and PT54466. Amino acid distances based on PT54465, red/green triangles are iron- and carbonate-coordinating amino acids.

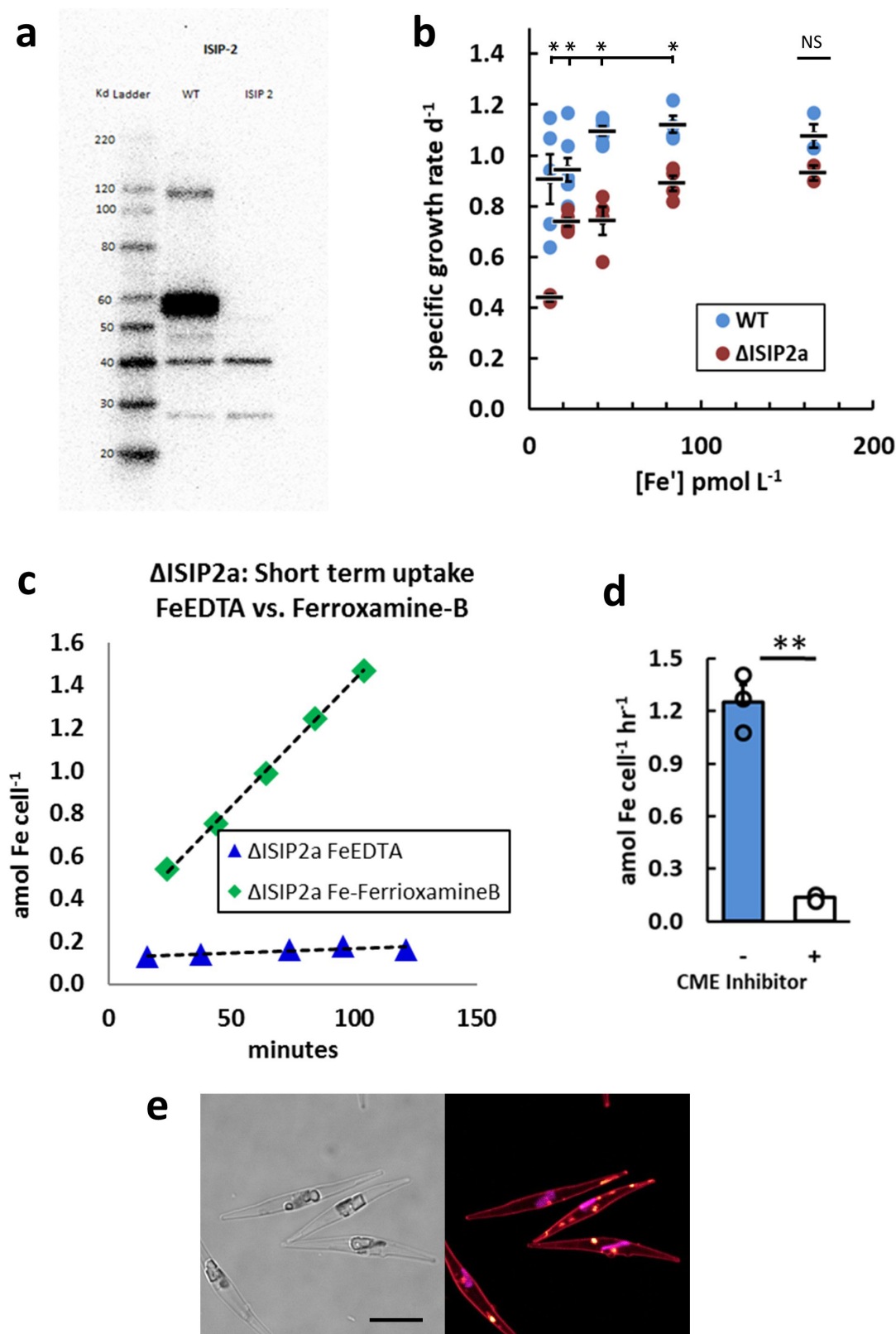


Extended Data Figure 2 | Estimated divergence times between transferrin, phytotransferrin and PBPs. Analyses were carried out using PhyloBayes (top) and BEAST (bottom). For the fossil calibration points used in generating the minimum and maximum constraints, see Supplementary Table 1. MYA, million years ago.



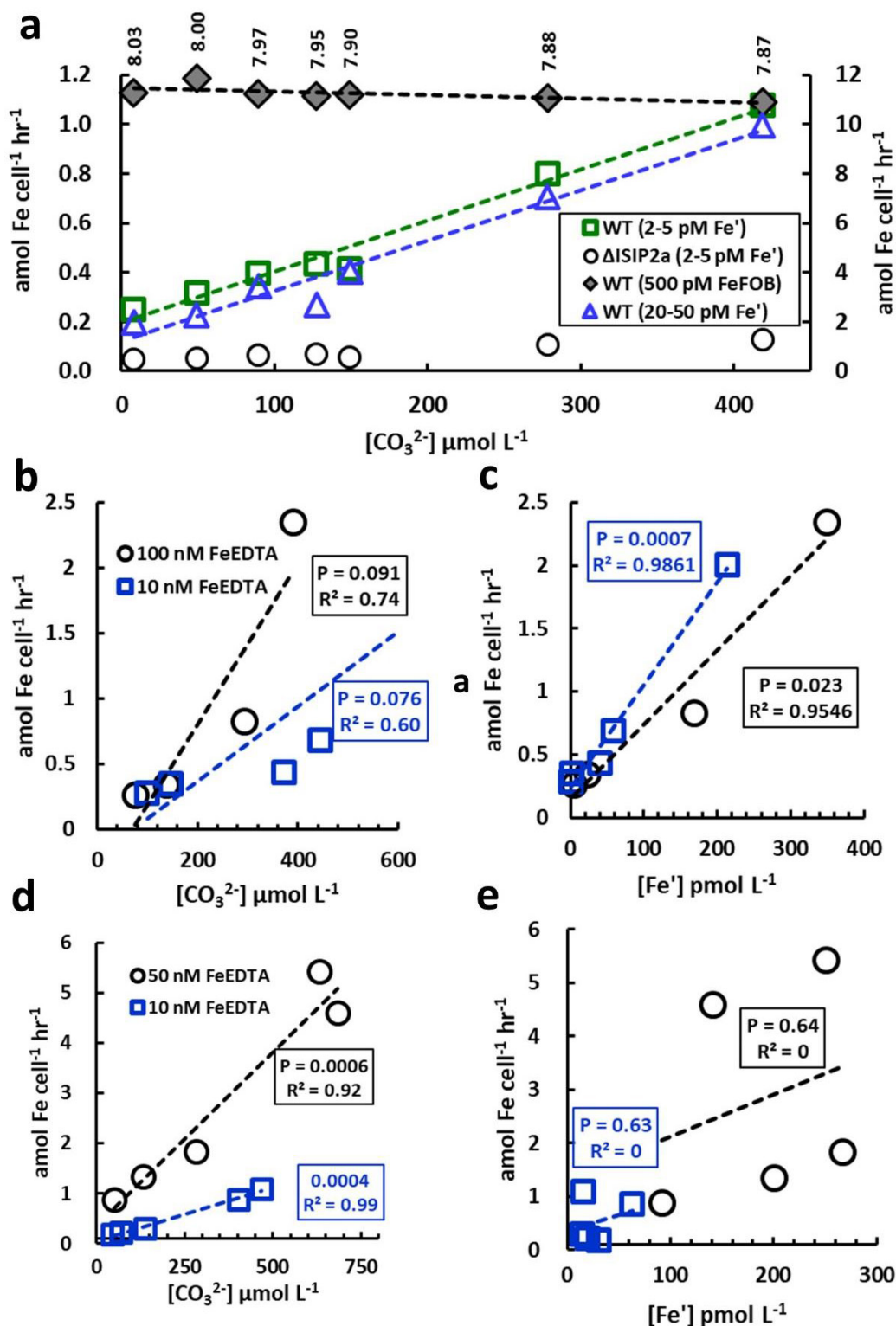
Extended Data Figure 3 | Bayesian (PhyloBayes) phylogenetic tree contrasted with alternate topology derived using maximum likelihood. *Prasinococcus capsulatus*, *Pyraminomonas obovata* and other chlorophyte

algae have copies of transferrin and phytotransferrin, whereas all other non-chlorophyte algae have only phytotransferrin. Branches are colour-coded by phylogenetic group. Scale bar, 0.5 substitutions per position.



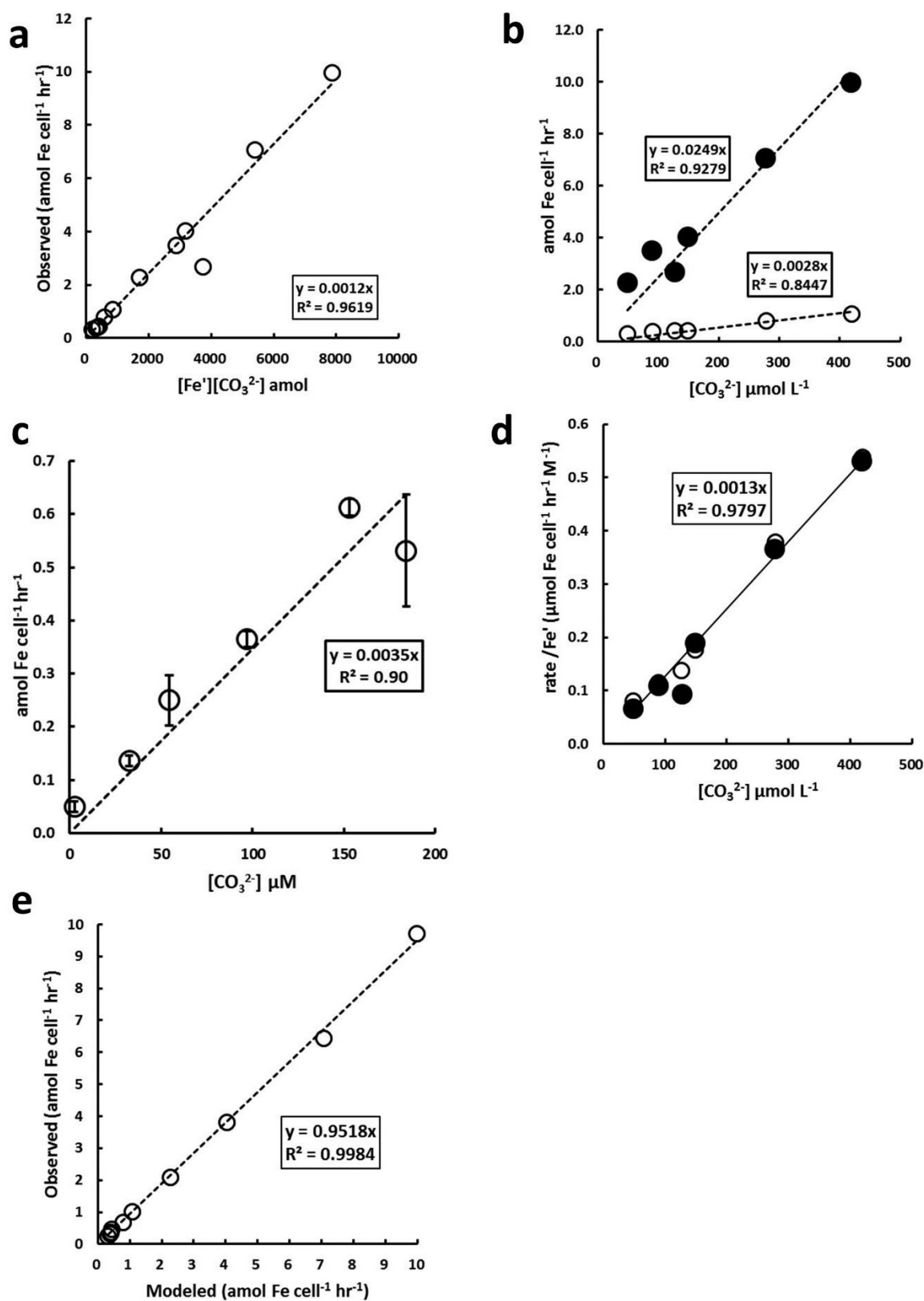
Extended Data Figure 4 | Characterization of ISIP2A. **a**, Western blot of wild-type *P. tricornutum* compared to *ISIP2A*-knockout strains (*ΔISIP2A*). The estimated mass of ISIP2A protein is 57 kDa. **b**, Specific growth rates of wild type compared to *ΔISIP2A*. **c**, Uptake of ^{59}Fe -ferrioxamine B is unaffected in *ΔISIP2A*, suggesting an alternate pathway for uptake of iron-siderophore complexes. **d**, Effect of a clathrin-mediated endocytosis (CME) inhibitor on short-term $^{59}Fe'$ uptake rates, wild-type *P. tricornutum*. **e**, Upon addition of iron to iron-limited cells, the

membrane-impermeable FM1-43 stain is internalized into vesicle-like inclusions. Scale bar, 5 μm . Pink is plastid auto-fluorescence. **b**, Data are mean \pm s.e.m.; two-sided heteroscedastic *t*-test, n = biological replicates for wild-type and *ISIP2A* are indicated in brackets as (*WT*, *ISIP2A*) and *P* values are indicated. 12 $pmol\ Fe'$ (5, 3), P = 0.008; 22 $pmol\ Fe'$ (7, 5), P = 0.003; 43 $pmol\ Fe'$ (6, 4), P = 0.006; 83 $pmol\ Fe'$ (4, 4), P = 0.007; 165 $pmol\ Fe'$ (3, 3), P = 0.07. **d**, Data are mean \pm s.e.m.; n = 3 biological replicates; two-sided *t*-test, P = 0.009.



Extended Data Figure 5 | Reconciliation of the carbonate effect with the influence of acidification on seawater iron chemistry. **a**, Linearized representation of Fig. 3a, with high iron uptake rates (blue triangles) plotted on the secondary y axis, and pH values for each $[\text{CO}_3^{2-}]$ listed at top. Fe-FOB uptake rates decrease with pH, consistent with the findings of ref. 26, whereas Fe' uptake rates increase with decreasing pH, inconsistent with the hypothesized effects of acidification on iron-EDTA chemistry²⁶. This inconsistency is resolved when uptake rates are plotted as a function

of the synergistic interaction between Fe' and CO_3^{2-} (Fig. 3b). **b**, **c**, Under CO_2 -induced acidification, the strong influence of pH on $[\text{Fe}']$ results in a significant correlation of uptake to $[\text{Fe}']$, although the interaction Fe' and CO_3^{2-} results in a better fit (Fig. 4a, solid line). **d**, **e**, When the change in $[\text{Fe}']$ is constrained relative to pH and $[\text{CO}_3^{2-}]$, uptake rates are positively correlated with $[\text{CO}_3^{2-}]$ and uncorrelated to $[\text{Fe}']$, revealing the influence of the carbonate ion on Fe' uptake rates. For statistical analyses of the linear regressions, see Extended Data Table 4.



Extended Data Figure 6 | Derivation of second-order and constitutive rate constants from *P. tricornutum* resuspended in NaHCO_3 -manipulated medium. **a**, Regression of the interaction product of Fe' and CO_3^{2-} . Regression excludes the two observations in which the medium was not supplemented with NaHCO_3 . **b**, ^{59}Fe uptake rates for treatments incubated with low (2–5 pM Fe' , open symbols) and high (20–50 pM Fe' ,

closed symbols) ^{59}Fe versus $[\text{CO}_3^{2-}]$. **c**, Demonstration of reproducibility at low $[\text{Fe}']$. Data are mean \pm s.e.m.; $n = 3$ biological replicates. **d**, Uptake rates normalized to $[\text{Fe}']$, plotted against $[\text{CO}_3^{2-}]$. The slope has units of $\text{mol Fe cell}^{-1} \text{h}^{-1} (\text{M Fe}')^{-1} (\text{M CO}_3^{2-})^{-1}$, equivalent to the pseudo-first-order uptake rate with respect to carbonate. **e**, Rates of uptake calculated as a function of $[\text{Fe}']$ and $[\text{CO}_3^{2-}]$ versus measured rates.

Extended Data Table 1 | Uptake rates compared to the measured (in bold) and derived concentrations of Fe' and carbonic acid species, in Aquil uptake medium

| | | Fe' uptake | [DIC] | [TA] | | [CO ₃ ²⁻] | [Fe'] | [Fe']x[CO ₃ ²⁻] | [EDTA] | [Fe] _{TOTAL}] | log K' _{dark} |
|----------------------------------------|---------------------------|-----------------------------------------------|-----------------------|-----------------------|------|----------------------------------|----------------------|----------------------------------------|----------------------|-------------------------|------------------------|
| Treatment | | amol Fe'·cell ⁻¹ ·hr ⁻¹ | μmol kg ⁻¹ | μmol kg ⁻¹ | pH | μmol L ⁻¹ | pmol L ⁻¹ | fmol ² L ⁻² | μmol L ⁻¹ | nM | 20°C |
| 10 nM FeEDTA <i>P. tricornutum</i> | 6 mM NaHCO ₃ | 1.076 | 5557 | - | 7.87 | 419 | 2.0 | 0.84 | 100 | 10.8 | -7.731 |
| | 4 mM NaHCO ₃ | 0.800 | 3647 | - | 7.88 | 278 | 2.1 | 0.59 | 100 | 10.9 | -7.716 |
| | 2 mM NaHCO ₃ | 0.412 | 1884 | - | 7.90 | 149 | 2.3 | 0.35 | 100 | 11.0 | -7.674 |
| | 1.5 mM NaHCO ₃ | 0.433 | 1437 | - | 7.95 | 127 | 3.1 | 0.40 | 100 | 11.0 | -7.545 |
| | 1 mM NaHCO ₃ | 0.396 | 970 | - | 7.97 | 90 | 3.4 | 0.31 | 100 | 10.8 | -7.499 |
| | 0.5 mM NaHCO ₃ | 0.318 | 505 | - | 8.00 | 49 | 4.0 | 0.20 | 100 | 10.9 | -7.436 |
| | unamended | 0.250 | 75 | - | 8.03 | 8 | 4.9 | 0.04 | 100 | 10.9 | -7.347 |
| Treatment | | Fe' uptake | [DIC] | [TA] | | [CO ₃ ²⁻] | [Fe'] | [Fe']x[CO ₃ ²⁻] | [EDTA] | [Fe] _{TOTAL}] | log K' _{dark} |
| Treatment | | amol Fe'·cell ⁻¹ ·hr ⁻¹ | μmol kg ⁻¹ | μmol kg ⁻¹ | pH | μmol L ⁻¹ | pmol L ⁻¹ | fmol ² L ⁻² | μmol L ⁻¹ | nM | 20°C |
| 100 nM FeEDTA <i>P. tricornutum</i> | 6 mM NaHCO ₃ | 9.967 | 5557 | - | 7.87 | 419 | 18.8 | 7.89 | 100 | 101.4 | -7.731 |
| | 4 mM NaHCO ₃ | 7.071 | 3647 | - | 7.88 | 278 | 19.4 | 5.42 | 100 | 101.1 | -7.716 |
| | 2 mM NaHCO ₃ | 4.032 | 1884 | - | 7.90 | 149 | 21.3 | 3.17 | 100 | 100.5 | -7.674 |
| | 1.5 mM NaHCO ₃ | 2.694 | 1437 | - | 7.95 | 127 | 29.1 | 3.71 | 100 | 102.0 | -7.545 |
| | 1 mM NaHCO ₃ | 3.433 | 970 | - | 7.97 | 90 | 32.1 | 2.88 | 100 | 101.2 | -7.499 |
| | 0.5 mM NaHCO ₃ | 2.264 | 505 | - | 8.00 | 49 | 34.7 | 1.71 | 100 | 94.7 | -7.436 |
| | unamended | 1.960 | 75 | - | 8.03 | 8 | 45.2 | 0.36 | 100 | 100.4 | -7.347 |
| Treatment | | Fe' uptake | [DIC] | [TA] | | [CO ₃ ²⁻] | [Fe'] | [Fe']x[CO ₃ ²⁻] | [EDTA] | [Fe] _{TOTAL}] | log K' _{dark} |
| Treatment | | amol Fe'·cell ⁻¹ ·hr ⁻¹ | μmol kg ⁻¹ | μmol kg ⁻¹ | pH | μmol L ⁻¹ | pmol L ⁻¹ | fmol ² L ⁻² | μmol L ⁻¹ | nM | 20°C |
| 10 nM FeEDTA <i>ΔISIP2a</i> | 6 mM NaHCO ₃ | 0.129 | 5557 | - | 7.87 | 419 | 2.0 | 0.84 | 100 | 10.8 | -7.731 |
| | 4 mM NaHCO ₃ | 0.107 | 3647 | - | 7.88 | 278 | 2.1 | 0.59 | 100 | 10.9 | -7.716 |
| | 2 mM NaHCO ₃ | 0.057 | 1884 | - | 7.90 | 149 | 2.3 | 0.34 | 100 | 10.9 | -7.674 |
| | 1.5 mM NaHCO ₃ | 0.070 | 1437 | - | 7.95 | 127 | 3.1 | 0.40 | 100 | 10.9 | -7.545 |
| | 1 mM NaHCO ₃ | 0.066 | 970 | - | 7.97 | 90 | 3.5 | 0.31 | 100 | 11.0 | -7.499 |
| | 0.5 mM NaHCO ₃ | 0.052 | 505 | - | 8.00 | 49 | 4.0 | 0.20 | 100 | 10.9 | -7.436 |
| | unamended | 0.049 | 75 | - | 8.03 | 8 | 4.9 | 0.04 | 100 | 10.9 | -7.347 |
| Treatment | | Fe' uptake | [DIC] | [TA] | | [CO ₃ ²⁻] | [FeFOB] | [Fe']x[CO ₃ ²⁻] | | | |
| Treatment | | amol Fe'·cell ⁻¹ ·hr ⁻¹ | μmol kg ⁻¹ | μmol kg ⁻¹ | pH | μmol L ⁻¹ | pmol L ⁻¹ | fmol ² L ⁻² | | | |
| 500 pM FeFOB <i>P. tricornutum</i> | 6 mM NaHCO ₃ | 1.090 | 5557 | - | 7.87 | 419 | 500 | - | | | |
| | 4 mM NaHCO ₃ | 1.106 | 3647 | - | 7.88 | 278 | 500 | - | | | |
| | 2 mM NaHCO ₃ | 1.117 | 1884 | - | 7.90 | 149 | 500 | - | | | |
| | 1.5 mM NaHCO ₃ | 1.116 | 1437 | - | 7.95 | 127 | 500 | - | | | |
| | 1 mM NaHCO ₃ | 1.123 | 970 | - | 7.97 | 90 | 500 | - | | | |
| | 0.5 mM NaHCO ₃ | 1.187 | 505 | - | 8.00 | 49 | 500 | - | | | |
| | unamended | 1.126 | 75 | - | 8.03 | 8 | 500 | - | | | |

These data were used to generate Fig. 3a, b. Note, total alkalinity (TA) was not measured in any of these samples because 2 mM Tris-HCl was used to maintain the pH at approximately 8.0.

Extended Data Table 2 | Uptake rates compared to measured (in bold) and derived concentrations for both Fe' and carbonic acid species

| | | CO ₂ (ppm) | Fe' uptake amol cell ⁻¹ hr ⁻¹ | [DIC] μmol kg ⁻¹ | [TA] μmol kg ⁻¹ | pH | [CO ₃ ²⁻] μmol L ⁻¹ | [Fe'] pmol L ⁻¹ | [Fe']x[CO ₃ ²⁻] amol ² L ⁻² | K' _{DARK} 21.5°C | [EDTA] μmol L ⁻¹ | [Fe:EDTA] nmol L ⁻¹ |
|-----------------------------------|-------------------------------|--------------------------|--------------------------------------------------------|--------------------------------|-------------------------------|--------------|----------------------------------------------------------|-------------------------------|-----------------------------------------------------------------------------|------------------------------|--------------------------------|-----------------------------------|
| Nat. Seawater vs. Aquil (Fig. 3d) | Natural Seawater Ross Sea | 200 | 0.483 | 1837 | 2328 | 8.36 | 320 | 21.1 | 6767 | -6.550 | 100 | 7.5 |
| | | amb. | 0.064 | 2099 | 2324 | 7.94 | 154 | 2.0 | 312 | -7.570 | 100 | 7.5 |
| | | 1000 | 0.101 | 2045 | 2342 | 8.07 | 199 | 4.2 | 832 | -7.254 | 100 | 7.5 |
| | Natural Seawater North Pac | 200 | 0.645 | 1874 | 2387 | 8.38 | 337 | 23.6 | 7952 | -6.502 | 100 | 7.5 |
| | | amb. | 0.175 | 2058 | 2371 | 8.09 | 210 | 4.7 | 979 | -7.206 | 100 | 7.5 |
| | | 1000 | 0.128 | 2091 | 2397 | 8.07 | 206 | 4.2 | 860 | -7.254 | 100 | 7.5 |
| | Synthetic Seawater (Aquil) | 200 | 0.525 | 509 | 1012 | 8.81 | 189 | 26.1 | 4943 | -5.458 | 100 | 7.5 |
| | | amb. | 0.315 | 2070 | 2450 | 8.18 | 255 | 7.7 | 1973 | -6.987 | 100 | 7.5 |
| | | 1000 | 0.092 | 830 | 1047 | 8.14 | 94 | 6.2 | 581 | -7.084 | 100 | 7.5 |
| Subs. Anions (Fig. 3c) | *Unamended | | 0.049 | 27 | - | 8.059 | 2.6 | 5.2 | 13.5 | -7.282 | 100 | 7.5 |
| | *Carbonate | | 0.263 | 1149 | - | 8.003 | 97.2 | 3.8 | 278.5 | -7.418 | 100 | 7.5 |
| | Formate | | 0.052 | 27 | - | 8.063 | 2.6 | 5.4 | 13.9 | -7.271 | 100 | 7.5 |
| | Acetate | | 0.090 | 41 | - | 8.051 | 3.9 | 5.0 | 19.5 | -7.299 | 100 | 7.5 |
| | Oxalate | | 0.060 | 28 | - | 8.059 | 2.8 | 5.2 | 14.8 | -7.280 | 100 | 7.5 |
| | Glycolate | | 0.097 | 20 | - | 8.067 | 2.2 | 5.5 | 11.7 | -7.263 | 100 | 7.5 |
| | Pyruvate | | 0.099 | 36 | - | 8.019 | 3.3 | 4.2 | 13.7 | -7.377 | 100 | 7.5 |
| | | | | | | | | | | | | |
| NaHCO ₃ (Fig. 6c) | *Unamended | | 0.049 | 27 | - | 8.059 | 2.6 | 5.2 | 10.1 | -7.282 | 100 | 7.5 |
| | 0.5 mM NaHCO ₃ | | 0.096 | 358 | - | 8.036 | 32.6 | 4.6 | 112.7 | -7.337 | 100 | 7.5 |
| | 1.0 mM NaHCO ₃ | | 0.170 | 603 | - | 8.036 | 54.5 | 4.6 | 250.9 | -7.336 | 100 | 7.5 |
| | *1.5 mM NaHCO ₃ | | 0.263 | 1149 | - | 8.003 | 97.2 | 3.8 | 278.5 | -7.418 | 100 | 7.5 |
| | 2.0 mM NaHCO ₃ | | 0.318 | 1806 | - | 8.004 | 153.1 | 3.8 | 441.4 | -7.415 | 100 | 7.5 |
| | 2.5 mM NaHCO ₃ | | 0.420 | 2423 | - | 7.953 | 184.1 | 2.9 | 398.9 | -7.539 | 100 | 7.5 |

* duplicate data used in both experiments

Top, uptake rates of cells resuspended in two types of natural compared to synthetic seawater at ambient or modified with 200 and 1,000 p.p.m. CO₂. Uptake rates are an mean of three biological replicates. These data were used to generate Fig. 3d. Bottom, uptake rates of cells resuspended in synthetic seawater (Aquil medium) supplemented with 100 μM of carbonate or alternate synergistic anions, and data show the reproducibility of the NaHCO₃ manipulation experiment. All uptake rates are an average of three biological replicates. These data were used to generate Fig. 3c and Extended Data Fig. 6c, respectively.

Extended Data Table 3 | Measured (in bold) and derived values for CO₂ and pH manipulation experiments

| | | CO ₂ | Fe' uptake | [DIC] | [TA] | | [CO ₃ ²⁻] | [Fe'] | [Fe']x[CO ₃ ²⁻] | [EDTA] | [Fe _{TOTAL}] | log K' _{DARK} | | Fv/Fm |
|---------------------------------------------------------|---------------|-----------------|------------------------------------------|-----------------------|-----------------------|------|----------------------------------|----------------------|----------------------------------------|----------------------|------------------------|------------------------|------|-------|
| | | TREATMENT | amol cell ⁻¹ hr ⁻¹ | μmol kg ⁻¹ | μmol kg ⁻¹ | pH | μmol L ⁻¹ | pmol L ⁻¹ | amol ² L ⁻² | μmol L ⁻¹ | nM | 20°C | μ | |
| pH/CO ₂ manipulations Δ[Fe']unconstrained | 10 nM FeEDTA | 50 ppm | 2.009 | 2052 | 2937 | 8.66 | 605 | 214 | 129.3 | 100 | 14.4 | -5.829 | 1.21 | 0.606 |
| | | 200 ppm | 0.684 | 1903 | 2583 | 8.52 | 444 | 58 | 25.7 | 100 | 8.5 | -6.164 | 1.02 | 0.613 |
| | | 400 ppm | 0.434 | 1802 | 2392 | 8.45 | 373 | 40 | 14.9 | 100 | 8.4 | -6.325 | 1.13 | 0.577 |
| | | 1000 ppm | 0.349 | 2162 | 2391 | 7.90 | 146 | 2 | 0.3 | 100 | 8.5 | -7.664 | 0.72 | 0.513 |
| | | 5000 ppm | 0.280 | 2488 | 2623 | 7.67 | 100 | 0.8 | 0.1 | 100 | 14.2 | -8.237 | 0.48 | 0.431 |
| | 100 nM FeEDTA | 50 ppm | 7.358 | 1979 | 2891 | 8.70 | 620 | 1831§ | 1135.3 | 100 | 100.0 | -5.737 | 1.14 | 0.587 |
| | | 200 ppm | 2.348 | 2000 | 2596 | 8.40 | 391 | 350 | 136.5 | 100 | 100.9 | -6.460 | 1.22 | 0.605 |
| | | 400 ppm | 0.832 | 1989 | 2446 | 8.27 | 292 | 168 | 49.0 | 100 | 97.5 | -6.764 | 1.29 | 0.611 |
| | | 1000 ppm | 0.345 | 2092 | 2311 | 7.89 | 138 | 22 | 3.1 | 100 | 109.4 | -7.691 | 0.93 | 0.592 |
| | | 5000 ppm | 0.267 | 2288 | 2386 | 7.58 | 76 | 3 | 0.3 | 100 | 97.5 | -8.446 | 0.70 | 0.504 |

§The calculated concentration of Fe' exceeds the 750 pmol Fe' L⁻¹ iron hydroxide precipitation limit, therefore this datapoint was not used in data models

| | | CO ₂ | Fe' uptake | [DIC] | [TA] | | [CO ₃ ²⁻] | [Fe'] | [Fe']x[CO ₃ ²⁻] | [EDTA] | [Fe _{TOTAL}] | log K' _{DARK} | | Fv/Fm |
|-------------------------------------------------------|--------------|-----------------|------------------------------------------|-----------------------|-----------------------|------|----------------------------------|----------------------|----------------------------------------|----------------------|------------------------|------------------------|------|-------|
| | | TREATMENT | amol cell ⁻¹ hr ⁻¹ | μmol kg ⁻¹ | μmol kg ⁻¹ | pH | μmol L ⁻¹ | pmol L ⁻¹ | amol ² L ⁻² | μmol L ⁻¹ | nM | 20°C | μ | |
| pH/CO ₂ manipulations Δ[Fe']constrained | 10 nM FeEDTA | 50 ppm | 1.089 | 1587 | 2335 | 8.63 | 468 | 16 | 8.4 | 1000 | 12.0 | -5.888 | 1.05 | 0.584 |
| | | 200 ppm | 0.865 | 1803 | 2434 | 8.47 | 403 | 63 | 25.5 | 100 | 12.1 | -6.281 | 1.11 | 0.598 |
| | | 400 ppm | 0.301 | 1874 | 2103 | 7.92 | 138 | 14 | 1.9 | 20 | 12.0 | -7.630 | 1.02 | 0.578 |
| | | 1000 ppm | 0.221 | 2088 | 2183 | 7.56 | 69 | 19 | 1.3 | 2 | 12.2 | -8.499 | 0.74 | 0.512 |
| | | 5000 ppm | 0.182 | 2900 | 2874 | 7.23 | 46 | 30 | 1.4 | 0.2 | 11.7 | -9.286 | 0.77 | 0.517 |
| | 50 nM FeEDTA | 50 ppm | 4.605 | 1991 | 2987 | 8.75 | 683 | 141 | 96.0 | 1000 | 55.2 | -5.594 | 1.12 | 0.603 |
| | | 200 ppm | 5.432 | 2315 | 3210 | 8.61 | 631 | 250 | 157.8 | 250 | 55.2 | -5.946 | 1.23 | 0.617 |
| | | 400 ppm | 1.847 | 1632 | 2103 | 8.33 | 282 | 267 | 75.4 | 50 | 55.1 | -6.616 | 1.06 | 0.581 |
| | | 1000 ppm | 1.348 | 1987 | 2201 | 7.87 | 132 | 200 | 26.4 | 5 | 55.4 | -7.744 | 1.08 | 0.564 |
| | | 5000 ppm | 0.888 | 1918 | 1975 | 7.44 | 49 | 91 | 4.5 | 1 | 55.4 | -8.783 | 0.92 | 0.541 |

Top, CO₂ manipulations where the Fe:EDTA ratio was held constant, and Fe' concentrations were allowed to change with changes in pH. Bottom, CO₂ manipulations in which the pH was allowed to change whereas the Fe:EDTA ratio was manipulated, constraining changes in [Fe'] to a narrow range (for full statistical analysis, see Extended Data Table 4).

Extended Data Table 4 | Statistical analyses of NaHCO₃ and pH and CO₂ manipulations**Linear regression analysis, Fe' uptake**

| | FIGURE | Data set | term | DF | RSE | adj. R ² | F-stat | P-value | comment |
|----------------------------------------------------------|---------|------------------------|----------------------------------------|----|--------------------------|---------------------|--------|--------------------------|-----------------|
| NaHCO ₃ manipulation (ΔpH constrained) | 3a | WT Low Fe' | [CO ₃ ²⁻] | 5 | 4.97 × 10 ⁻¹⁷ | 0.972 | 212 | 2.76 × 10 ⁻⁵ | |
| | 3a | WT High Fe' | [CO ₃ ²⁻] | 5 | 6.13 × 10 ⁻¹⁶ | 0.957 | 135 | 8.29 × 10 ⁻⁵ | |
| | 3a | ΔISIP2a Low Fe' | [CO ₃ ²⁻] | 5 | 8.54 × 10 ⁻¹⁸ | 0.921 | 71 | 3.85 × 10 ⁻⁴ | |
| | | WT Low Fe' | [Fe'] | 5 | 1.86 × 10 ⁻¹⁶ | 0.611 | 10 | 2.32 × 10 ⁻² | |
| | | WT High Fe' | [Fe'] | 5 | 1.95 × 10 ⁻¹⁵ | 0.565 | 9 | 3.13 × 10 ⁻² | |
| | | ΔISIP2a Low Fe' | [Fe'] | 5 | 2.12 × 10 ⁻¹⁷ | 0.514 | 7 | 4.23 × 10 ⁻² | |
| | 3b | WT (Combined Low/High) | [Fe']x[CO ₃ ²⁻] | 12 | 5.96 × 10 ⁻¹⁶ | 0.957 | 292 | 8.70 × 10 ⁻¹⁰ | |
| CO ₂ /pH manipulation (ΔFe' unconstrained) | Ext. 5b | WT Low Fe' | [CO ₃ ²⁻] | 3 | 4.54 × 10 ⁻¹⁶ | 0.601 | 7 | 7.68 × 10 ⁻² | Not Significant |
| | Ext. 5b | WT High Fe' | [CO ₃ ²⁻] | 2 | 4.92 × 10 ⁻¹⁶ | 0.741 | 10 | 9.06 × 10 ⁻² | Not Significant |
| | Ext. 5c | WT Low Fe' | [Fe'] | 3 | 9.81 × 10 ⁻¹⁷ | 0.981 | 212 | 7.02 × 10 ⁻⁴ | |
| | Ext. 5c | WT High Fe' | [Fe'] | 2 | 2.52 × 10 ⁻¹⁶ | 0.932 | 42 | 2.30 × 10 ⁻² | |
| | 4a | WT (Combined Low/High) | [Fe']x[CO ₃ ²⁻] | 7 | 9.13 × 10 ⁻¹⁷ | 0.987 | 588 | 5.16 × 10 ⁻⁸ | Solid line |
| CO ₂ /pH manipulation (ΔFe' constrained) | Ext. 5d | WT Low Fe' | [CO ₃ ²⁻] | 3 | 4.85 × 10 ⁻¹⁷ | 0.987 | 292 | 4.36 × 10 ⁻⁴ | |
| | Ext. 5d | WT High Fe' | [CO ₃ ²⁻] | 3 | 5.74 × 10 ⁻¹⁶ | 0.922 | 48 | 6.15 × 10 ⁻⁴ | |
| | Ext. 5e | WT Low Fe' | [Fe'] | 3 | 4.60 × 10 ⁻¹⁶ | 0 | 0 | 6.34 × 10 ⁻¹ | Not Significant |
| | Ext. 5e | WT High Fe' | [Fe'] | 3 | 2.27 × 10 ⁻¹⁵ | 0 | 0 | 6.47 × 10 ⁻¹ | Not Significant |
| | 4b | WT (Combined Low/High) | [Fe']x[CO ₃ ²⁻] | 8 | 5.90 × 10 ⁻¹⁶ | 0.898 | 80 | 1.93 × 10 ⁻⁵ | Dashed line |

Significant interactions: 2-way Analysis of Variation (ANOVA), Fe' uptake

| FIGURE | Experiment type | term | DF | Mean Sq | F | p-value |
|-----------------------|--------------------------------------------------------|----------------------------------------|----|--------------------------|-----|-------------------------|
| 3b | NaHCO ₃ manipulation (pH constrained) | [Fe'] | 1 | 3.71 × 10 ⁻²⁹ | 131 | 4.55 × 10 ⁻⁷ |
| | | [CO ₃ ²⁻] | 1 | 4.13 × 10 ⁻²⁹ | 146 | 2.73 × 10 ⁻⁷ |
| | | [Fe']x[CO ₃ ²⁻] | 1 | 2.68 × 10 ⁻²⁹ | 95 | 2.02 × 10 ⁻⁶ |
| 4a (Solid Line) | pH/CO ₂ manipulation (Fe' unconstrained) | [Fe'] | 1 | 4.53 × 10 ⁻³⁰ | 450 | 4.33 × 10 ⁻⁶ |
| | | [CO ₃ ²⁻] | 1 | 1.10 × 10 ⁻³¹ | 11 | 2.14 × 10 ⁻² |
| | | [Fe']x[CO ₃ ²⁻] | 1 | 2.63 × 10 ⁻³¹ | 26 | 3.75 × 10 ⁻³ |
| 4a (Dash Line) | pH/CO ₂ manipulation (Fe' constrained) | [Fe'] | 1 | 1.38 × 10 ⁻²⁹ | 43 | 5.89 × 10 ⁻⁴ |
| | | [CO ₃ ²⁻] | 1 | 1.22 × 10 ⁻²⁹ | 38 | 8.10 × 10 ⁻⁴ |
| | | [Fe']x[CO ₃ ²⁻] | 1 | 2.76 × 10 ⁻³⁰ | 9 | 2.58 × 10 ⁻² |

The *cis*-regulatory dynamics of embryonic development at single-cell resolution

Darren A. Cusanovich^{1*}, James P. Reddington^{2*}, David A. Garfield^{2†*}, Riza M. Daza¹, Delasa Aghamirzaie¹, Raquel Marco-Ferreres², Hannah A. Pliner¹, Lena Christiansen³, Xiaojie Qiu¹, Frank J. Steemers³, Cole Trapnell¹, Jay Shendure^{1,4§} & Eileen E. M. Furlong^{2§}

Understanding how gene regulatory networks control the progressive restriction of cell fates is a long-standing challenge. Recent advances in measuring gene expression in single cells are providing new insights into lineage commitment. However, the regulatory events underlying these changes remain unclear. Here we investigate the dynamics of chromatin regulatory landscapes during embryogenesis at single-cell resolution. Using single-cell combinatorial indexing assay for transposase accessible chromatin with sequencing (sci-ATAC-seq)¹, we profiled chromatin accessibility in over 20,000 single nuclei from fixed *Drosophila melanogaster* embryos spanning three landmark embryonic stages: 2–4 h after egg laying (predominantly stage 5 blastoderm nuclei), when each embryo comprises around 6,000 multipotent cells; 6–8 h after egg laying (predominantly stage 10–11), to capture a midpoint in embryonic development when major lineages in the mesoderm and ectoderm are specified; and 10–12 h after egg laying (predominantly stage 13), when each of the embryo's more than 20,000 cells are undergoing terminal differentiation. Our results show that there is spatial heterogeneity in the accessibility of the regulatory genome before gastrulation, a feature that aligns with future cell fate, and that nuclei can be temporally ordered along developmental trajectories. During mid-embryogenesis, tissue granularity emerges such that individual cell types can be inferred by their chromatin accessibility while maintaining a signature of their germ layer of origin. Analysis of the data reveals overlapping usage of regulatory elements between cells of the endoderm and non-myogenic mesoderm, suggesting a common developmental program that is reminiscent of the mesendoderm lineage in other species^{2–4}. We identify 30,075 distal regulatory elements that exhibit tissue-specific accessibility. We validated the germ-layer specificity of a subset of these predicted enhancers in transgenic embryos, achieving an accuracy of 90%. Overall, our results demonstrate the power of shotgun single-cell profiling of embryos to resolve dynamic changes in the chromatin landscape during development, and to uncover the *cis*-regulatory programs of metazoan germ layers and cell types.

We adapted our sci-ATAC-seq protocol¹ to work with nuclei from formaldehyde-fixed *Drosophila* embryos and concurrently implemented optimizations to increase the sensitivity by roughly an order of magnitude. The nuclei processed from each developmental time point were derived from hundreds of embryos of both sexes, and were therefore likely to include intermediate developmental states. Of 431 million sequenced read pairs, 70% mapped to the nuclear reference genome and were assigned a cell barcode (Extended Data Fig. 1a, b). Altogether, we obtained chromatin accessibility profiles for 23,085 cells across the three time points (12,904 ± 10,979 (mean ± s.d.) reads per cell after

de-duplication; minimum of 500 unique reads per cell (Extended Data Fig. 1c)). Sequenced fragments exhibited nucleosomal banding and were strongly enriched in DNase-hypersensitive sites (DHS) that have been defined in bulk *Drosophila* embryos⁵ (Extended Data Fig. 1d).

We partitioned the genome into 2-kb windows and scored each cell by whether any reads were observed in each window. For each time point, we performed latent semantic indexing¹ (LSI) using the 20,000 most frequently accessible windows and discarding the 10% of cells with the fewest reads. Of the 20,000 windows, 14,295 were common across all three time points (Extended Data Fig. 1e). Although measurements of accessibility in individual cells are naturally sparse (as there are only 2–4 genome equivalents per nucleus), the data are sufficiently structured to reveal subsets of cells with similar chromatin accessibility (Fig. 1a–c). To map the underlying regulatory elements, we aggregated data from cells within each of the largest 4–5 clades per time point to call peaks and summits of accessibility for each ‘*in silico*-sorted’ clade (Fig. 1d). Merging summits across all time points and clades identified 53,133 potential *cis*-regulatory elements, 40,967 of which have clade-specific accessibility in at least one time point (Supplementary Table 1); including 12,605 at 2–4 h, 25,615 at 6–8 h and 28,253 at 10–12 h after egg laying (Extended Data Fig. 1f). These results reveal the highly dynamic and heterogeneous nature of chromatin accessibility during embryogenesis, with roughly twice as many differentially accessible sites identified at the later time points compared with the earlier one.

To determine the identity of each cell clade, we compared accessible regions to 3,841 developmental enhancers^{6–8} and 9,356 gene promoters^{9,10} with characterized tissue activity across embryogenesis. The enrichments of clade-specific promoter-distal (putative enhancers) and promoter-proximal (putative promoters) elements gave consistent results (Supplementary Table 2). The four major clades at 6–8 h and 10–12 h correspond to the three major germ layers, with two subdivisions: ectoderm, which is split into neurogenic (clade 1) and non-neurogenic (clade 2) lineages, and mesoderm, which is split into myogenic mesoderm (clade 3) and non-myogenic mesoderm (such as fat body and haemocytes) combined with endoderm (clade 4) (Extended Data Fig. 2, Supplementary Table 2). The latter indicates that non-myogenic mesoderm and endoderm exhibit similar chromatin accessibility, suggesting a shared developmental program. Although, to our knowledge, *Drosophila* mesoderm and endoderm have not been shown to share a common origin, this is highly reminiscent of the mesendoderm lineage in *Caenorhabditis elegans*², sea urchins³ and vertebrates⁴. Of the 53,133 potential *cis*-regulatory elements, 35,963 are distal (putative enhancers); 12% overlap characterized developmental enhancers and 48% overlap putative enhancers identified from bulk DHS data⁵ (based on 1-bp overlap). Conversely, of the

¹Department of Genome Sciences, University of Washington, Seattle, Washington, USA. ²European Molecular Biology Laboratory (EMBL), Genome Biology Unit, Heidelberg, Germany. ³illumina, San Diego, California, USA. ⁴Howard Hughes Medical Institute, Seattle, Washington, USA. [†]Present address: IRI Life Sciences, Humboldt Universität zu Berlin, Berlin, Germany.

*These authors contributed equally to this work.

§These authors jointly supervised this work.

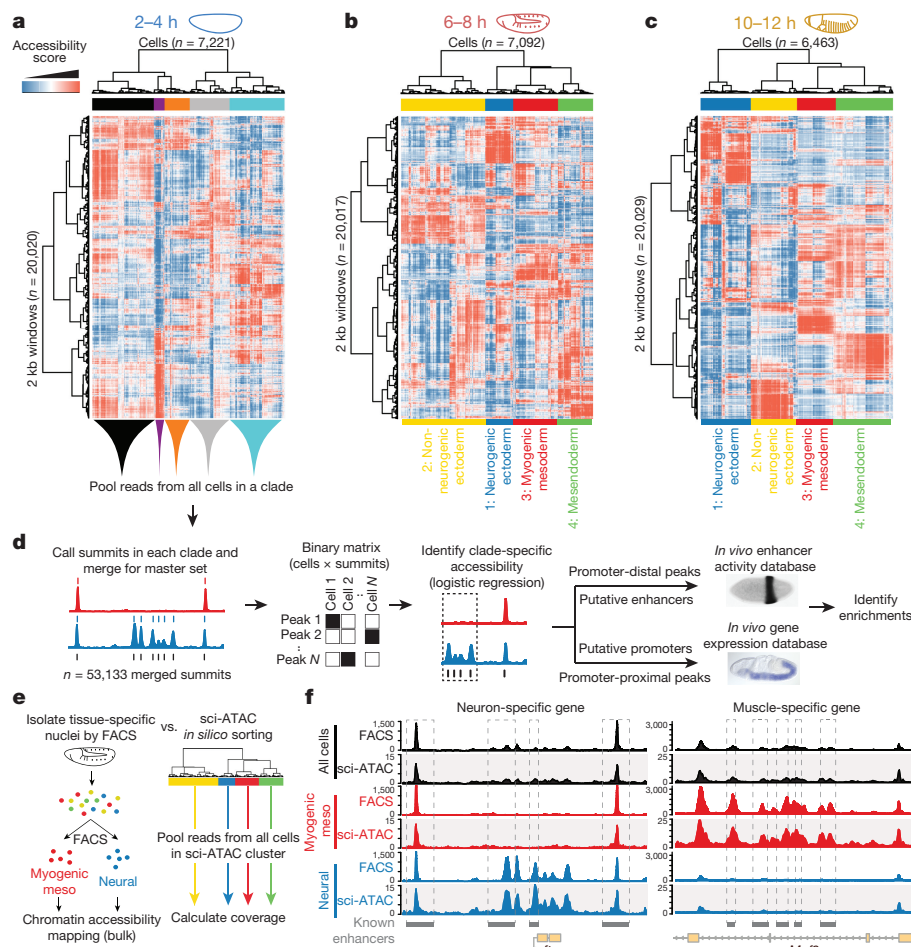


Figure 1 | Single-cell profiling of chromatin accessibility across *Drosophila* embryogenesis. **a–c**, Heat maps of binarized, LSI-transformed, clustered read counts for single cells (columns) in 2-kb windows across the genome (rows) at 2–4 h (**a**), 6–8 h (**b**) and 10–12 h (**c**) after egg laying. Major clades are assignable to germ layers at post-gastrulation time points (**b**, **c**). **d**, Approach to annotation of clades by intersecting clade-specific peaks of chromatin accessibility with enhancer activity and gene expression. *In situ* image of enhancer activity (black stain) from ref. 7; RNA *in situ* (blue stain) from the Berkeley *Drosophila* Genome Project^{10,31,32}. **e**, Comparing fluorescence-activated cell sorting combined with DNase I sequencing (FACS–DNase-seq) and *in silico*

sorting with sci-ATAC-seq. Nuclei from myogenic mesoderm and neurons were isolated from 6–8-h embryos using antibodies against tissue-specific regulatory proteins Mef2 (myogenic mesoderm) and Elav (neurons), sorted by FACS and analysed by DNase-seq. *In silico* sorts from sci-ATAC-seq were built by pooling reads from all cells within each LSI-defined clade. **f**, Library-size-normalized coverage tracks from FACS–DNase-seq (top graph for each clade) and sci-ATAC-seq *in silico* sorts (bottom graph for each clade) for whole embryo (black), mesodermal (red), and neuronal (blue) at 6–8 h. Shown are *ftz* (neuronal; left) and *Mef2* (mesodermal; right) loci. Known enhancers for each tissue are indicated.

3,841 characterized developmental enhancers, 2,533 (66%) overlapped regions of accessible chromatin identified in this study.

To validate *in silico* sorting and clade assignments, we used fluorescence-activated cell sorting (FACS) to isolate myogenic mesoderm and neuronal nuclei from 6–8 h embryos¹¹ to approximately 98% purity. Sorted nuclei were subjected to DNase I-hypersensitive-site sequencing (DNase-seq) in bulk, and the resulting accessibility maps were compared to our *in silico*-sorted (that is, clade-defined) sci-ATAC-seq data from 6–8-h embryos (Fig. 1e). The comparison shows notable similarity both globally (Spearman's $\rho > 0.85$ for matched versus 0.53 for non-matched comparisons) and at individual loci. For example, both methods show that previously characterized neuronal enhancers near the *ftz* gene are accessible in neurogenic ectoderm but not in myogenic mesoderm (Fig. 1f, left) and, conversely, that muscle enhancers of *Mef2* are accessible in myogenic mesoderm but not in neurogenic ectoderm (Fig. 1f, right).

The clade assignments are further supported by motif enrichments for transcription factor binding sites and transcription factor occupancy at putative enhancers. For example, at mid and late embryogenesis, motifs for the lineage-specifying factors Krüppel (Kr), tramtrack (Ttk) and runt (Run) were among the most enriched in neurogenic

ectoderm¹² (clade 1), Mef2 and Cf2 motifs were enriched in myogenic mesoderm¹³ (clade 3) and GATA motifs were enriched in mesoderm (clade 4) (Extended Data Fig. 3a–c, Supplementary Table 3). The presence of GATA motifs may reflect the conserved role of GATA factors in the specification of both non-myogenic mesoderm¹⁴ and endoderm¹⁵. Similarly, regions occupied by transcription factors with more constitutive roles, such as CTCF, exhibit similar accessibility across all clades (Extended Data Fig. 3d–g), whereas regions bound by myogenic transcription factors are more accessible in the myogenic mesodermal clade¹⁶ (Extended Data Fig. 3h–l).

Cells examined at 2–4 h after egg laying fall into five major clades (Fig. 1a) in which regulatory identities are clearly distinct from later stages in embryogenesis (Extended Data Fig. 4, Supplementary Table 2). The 2–4-h nuclei span embryos from the syncytial blastoderm, cellularization, gastrulation and early germ-band extension (stages 5–8), with the majority of embryos being pre-gastrulation (stage 5). Developmental transitions during these stages are very rapid, with cellularization (stage 5) lasting 40 min and onset of gastrulation (stage 6) lasting only 10 min. To capture finer granularity across these dynamic transitions, we applied *t*-distributed stochastic neighbour embedding (*t*-SNE)¹⁷ to the binary sci-ATAC-seq matrix of cells versus summits of

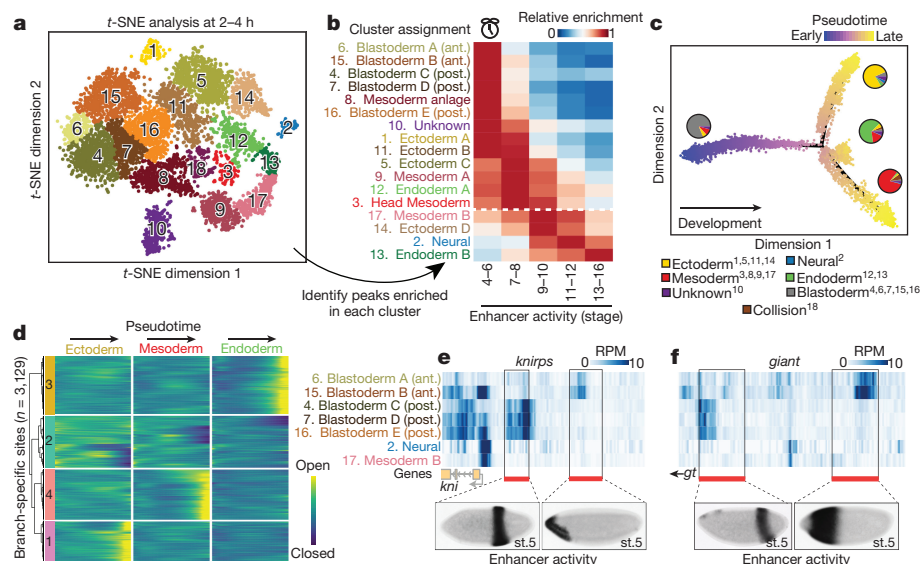


Figure 2 | Temporal dynamics and spatial heterogeneity in chromatin accessibility in the early embryo. **a**, *t*-SNE analysis of cells at 2–4 h. Clusters were defined by a density peak clustering algorithm (see Methods) and annotated on the basis of overlaps between cluster-enriched peaks and known tissue-specific enhancers or genes. **b**, Relative enrichment of enhancers that are active at different developmental stages in each cluster. Clusters below the white dashed line are likely to be derived from embryos outside the 2–4-h window, owing to female holding of older embryos. Ant., anterior; post., posterior. **c**, Pseudotime ordering of cells along a developmental trajectory. Cells were ordered in three dimensions (only two are shown) with DDRTree. Point colours correspond to cells' accessibility. Because of confounding differences in sex chromosome copy number between male and female nuclei (Extended Data Fig. 5), we restricted the matrix to autosomal elements.

accessibility. Because of confounding differences in sex chromosome copy number between male and female nuclei (Extended Data Fig. 5), we restricted the matrix to autosomal elements.

Density-peak clustering¹⁸ of cells after *t*-SNE enabled identification of 18 cell clusters at 2–4 h (Fig. 2a). Analysis of the relative enrichment of these clusters for active enhancers and transcription-factor occupancy (Supplementary Tables 4, 5) revealed marked differences in their developmental stages (Fig. 2b), highlighting developmental time as a major axis of variation within this time point. Notably, two of the developmentally early clusters were sex-biased (cluster 10: 85% male; cluster 1: 69% female). Whereas the identity of the male-biased cluster remains unclear, the female-biased cluster is enriched for enhancers that are active in brain anlage.

To evaluate this temporal ordering more formally, we used a graph-based method to arrange single cells into a developmental trajectory¹⁹. This 'pseudotemporal' ordering agreed well with the observed enrichments in cell clusters for active enhancers (Extended Data Fig. 6a–c). Notably, the trajectory split cells into three major branches that were consistent with our annotations of the major germ layers (neural cells are rare at this time point, as expected) (Fig. 2c). Pseudotemporal ordering also enabled us to explore the dynamics of sites that open or close within the 2–4-h window. We identified 12,165 sites with significant pseudotime-dependent temporal changes (1% false discovery rate (FDR)). Using a simple heuristic, we classified 5,219 (43%) of these sites as closing as pseudotime progressed; 5,133 (42%) as opening; and the remaining 1,813 (15%) as having more complex dynamics (Extended Data Fig. 6d–i, Supplementary Table 6). Many of the most pronounced changes match expectations, falling within gene loci that have dynamic roles during early embryogenesis. For example, the most significant closing site (P value = 5×10^{-224}) is within the *slam* locus, a gene that is essential for blastoderm cellularization during a very brief temporal window²⁰ (Extended Data Fig. 6g).

To identify sites that open or close specifically within individual germ-layer trajectories, we tested for pseudotime-dependent changes along each of the three paths (Fig. 2c) independently (with the potential

progression along the trajectory. Pie charts indicate relative frequencies of germ-layer assignments for cells in each branch. Superscript numbers in the key indicate which clusters from **a** were included in each category. **d**, Heat map of smoothed accessibility curves fit to sites (rows) for 100 bins of cells progressing through pseudotime (columns). Sites were clustered into four groups on the basis of their temporal dynamics. Only sites classified as branch-specific are shown. **e**, **f**, Heat maps of library-size-normalized read counts in the vicinity of the gap genes *knirps* (**e**) and *giant* (**f**). In each case, one characterized enhancer is known to drive anterior expression and the other drives posterior expression in blastoderm embryos (stage 5). *In situ* images of enhancer activity obtained from ref. 7.

caveat that these branches may be contaminated to some degree by cells from older embryos, owing to female 'holding'). This test identified 3,129 sites that were significantly pseudotime-dependent in only one branch, with 992, 1,071, and 1,066 restricted to the ectoderm, mesoderm and endoderm, respectively (Fig. 2d, Supplementary Tables 7–10). As with the global pseudotime ordering, sites associated with lineage-specific pseudotime exhibited dynamics consistent with biological expectation (for example, sites in the *heartless* (*hlt*)²¹, *GATAe*²², and *dachsous* (*ds*)²³ loci are accessible specifically in mesoderm, endoderm and ectoderm, respectively; Extended Data Fig. 6j–l).

Therefore, germ layers appear late in pseudotime at 2–4 h (Fig. 2c), yet developmentally early nuclei in this same window (as defined in Fig. 2b; clusters 6, 15, 4, 7, 8, 16) exhibit heterogeneous chromatin accessibility that reflects enhancer activity in refined spatial domains along the embryo's antero-posterior (A–P) and dorso-ventral (D–V) axes (Supplementary Table 5). For example, chromatin accessibility surrounding two gap genes, *knirps* (*kni*) and *giant* (*gt*), varies among developmentally early clusters (Fig. 2e, f). The expression of *knirps* and *giant* is spatially patterned in two broad stripes along the A–P axis of the embryo, each controlled by two enhancers driving either the posterior or the anterior expression⁷. The anterior enhancers of both genes have greater accessibility in cells of the presumptive anterior blastoderm clusters (clusters 6 and 15), while the posterior enhancers exhibit greater accessibility in the presumptive posterior blastoderm clusters (clusters 4, 7, and 16) (Fig. 2e, f). This example illustrates how despite being untargeted, sci-ATAC-seq can identify regulatory regions that are specifically accessible in spatially refined subsets of cells without the need for FACS sorting. Classic lineage-tracing and transplantation experiments showed that the broad fate and developmental potential of cells are largely determined at the cellular blastoderm stage, leading to the concept of a blastoderm fate map²⁴. Our data support the view that these early pre-gastrulation cell specification events are underpinned by spatial heterogeneity in chromatin accessibility.

Applying *t*-SNE to the later time points, during lineage commitment (6–8 h) and differentiation (10–12 h), revealed a fine-grained map of

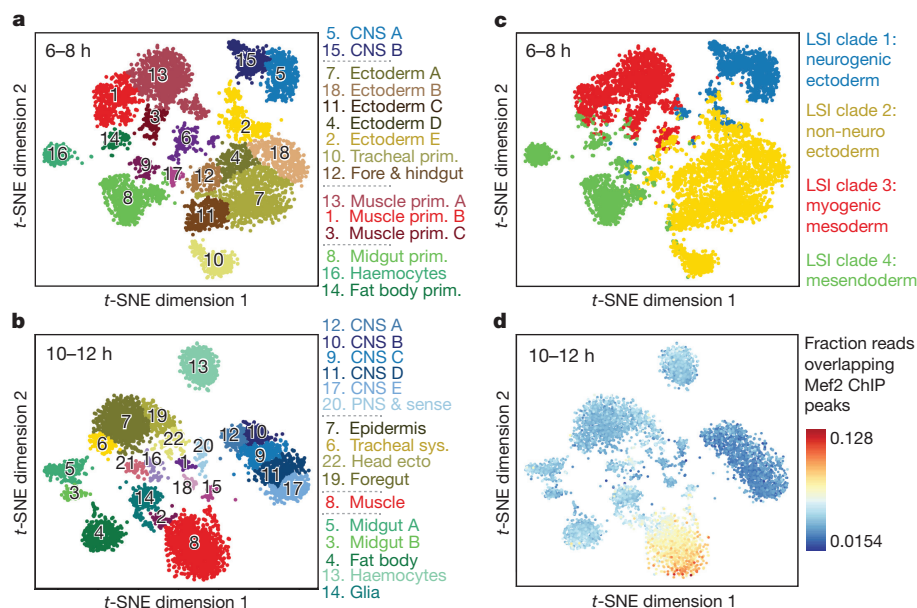


Figure 3 | Single cells are readily assigned to tissues and cell types based on chromatin accessibility. **a, b**, Clustering of sci-ATAC-seq data from the 6–8-h (**a**) and 10–12-h (**b**) time points after *t*-SNE dimensionality reduction. Clusters were annotated based on overlaps between cluster-enriched peaks and enhancers or genes with known tissue-specific activity. Three 6–8-h (6, 9, 17) and six 10–12-h (1, 2, 15, 16, 18, 21) clusters are likely to comprise multi-cell collisions based on library complexity and the distribution of reads mapping to the X chromosome (Extended Data Fig. 7). **c**, The 6–8-h *t*-SNE shown in **a**, coloured according to the original germ-layer assignment. **d**, The 10–12-h *t*-SNE shown in **b**, coloured according to the fraction of reads falling in Mef2 ChIP-seq peaks.

cell clusters that could be readily assigned to specific tissues or cell types (Fig. 3a, b; Supplementary Table 4). A few small clusters were identified as likely ‘collisions’ resulting from the combinatorial indexing, and were therefore discarded (purple clusters in Fig. 3a, b, Extended Data Fig. 7). For all remaining clusters, the cell-type assignments are broadly consistent with the germ-layer clade assignments (Fig. 3c, Extended Data Fig. 8), but with much finer granularity, whether we use information from either enhancer or gene-activity databases (Extended Data Fig. 9). For example, mesendoderm (clade 4 in Figs 1, 3c) is resolved into three separate clusters at 6–8 h, comprising the fat body (cluster 14) and haemocytes (cluster 16) from the non-myogenic mesoderm, and midgut (cluster 8) from the endoderm (Fig. 3a). Although we are clearly undersampling the number of cells present at these stages, the data are not obviously biased towards any particular tissue or cell type. The clusters’ tissue identities also match transcription factor occupancy by tissue-specific factors (Supplementary Table 4). For example, cells in cluster 8 (muscle) at 10–12 h are enriched for reads that overlap chromatin immunoprecipitation (ChIP) peaks for the key myogenic factor Mef2 at 10–12 h (Fig. 3d).

A major advantage of profiling chromatin accessibility is its potential to identify distal regulatory elements that shape gene expression. To determine whether elements that exhibit tissue-specific chromatin accessibility corresponded to bona fide tissue-specific enhancers, we tested 31 elements in transgenic embryos. We selected promoter-distal elements exhibiting clade-specific accessibility at 6–8 h and/or 10–12 h that did not overlap with previously characterized enhancers (Supplementary Table 11). No other criteria were used to bias the selection towards different classes of distal regulation (for example, enhancers versus insulators). Each putative regulatory element was cloned upstream of a minimal promoter driving a *lacZ* reporter and stably integrated into a common location in the *Drosophila* genome to minimize positional effects. Enhancer activity was then assessed across all stages of embryogenesis by *in situ* hybridization.

Notably, given the simple selection strategy, 94% (29 of 31) of tested regions functioned as developmental enhancers *in vivo* (Fig. 4, Extended Data Fig. 10, Supplementary Table 11). Furthermore, 90% (26 of 29) of active enhancers showed activity in the predicted tissue, with 23 being exclusive to that tissue (Extended Data Fig. 10, Supplementary Table 11). For example, elements specifically accessible in the neuronal, ectodermal or muscle clades show enhancer activity in the developing central nervous system (with some amnioserosa) (Fig. 4a), epidermis (Fig. 4b) and muscle (Fig. 4c), respectively. Elements that are specifically accessible in the mesendoderm clade

often act as enhancers in either the gut endoderm or haemocytes (mesoderm). Enhancer 4, for example, is accessible in cells of the developing midgut (endoderm) at both 6–8 h and 10–12 h, matching its activity in the anterior–posterior midgut during these stages (Fig. 4d). The only exceptions to our predictions were three of the seven elements that are specifically accessible in clade 4, which when tested were active in yolk nuclei (Extended Data Fig. 10). As the yolk is extra-embryonic, this was unexpected, and suggests a potential regulatory link between the yolk and mesendodermal tissues, which is supported by the role of the GATA transcription factor *serpent* in both yolk²⁵ and non-myogenic mesoderm¹⁴.

In summary, our results demonstrate the power of sci-ATAC-seq to not only elucidate the developmental dynamics of chromatin accessibility, but also for the large-scale prediction of *in vivo* enhancer activity. Altogether, we identified 30,075 putative distal regulatory elements exhibiting clade-specific accessibility (Supplementary Table 1). By combining reads from cells within each *t*-SNE cluster, we generated cell-type-specific tracks of chromatin accessibility, which reveal a wealth of differences between cell types, and a powerful resource for future investigations (<http://shiny.furlonglab.embl.de/scATACseqBrowser/>). We also provide site-by-cell matrices and vignettes to facilitate further exploration of the data (<http://atlas.gs.washington.edu>).

The sparsity of data from single-cell molecular profiling technologies, including sci-ATAC-seq, remains a challenge. Although insights can be derived by aggregating observations across subsets of cells, as done here, increasing the number of reads per cell will increase the granularity at which chromatin accessibility can be explored. Combinatorial indexing is subject to collisions; with our current strategy, around 12% of cell barcodes are expected to represent aggregates of two or more cells. Analogous to doublets in emulsion-based single-cell RNA sequencing, collisions primarily add noise to the aggregate profiles of clades, but can sometimes lead to artefactual clusters. We present a strategy for identifying such clusters here; however, collisions are likely to be more effectively overcome by additional rounds of combinatorial indexing²⁶, which would also increase throughput.

Looking forward, an expanded dataset that includes many more cells per time point and covers the entirety of *Drosophila* development has the potential to identify rarer cell types and reveal a fully continuous view of the landscape of chromatin accessibility as it unfolds. Our ability to understand how changes in the regulatory landscape underlie lineage commitment would be greatly aided by the concurrent measurement of chromatin accessibility and transcription. In the long term, the integration of chromatin state, transcriptional output²⁶, lineage history^{27,28}

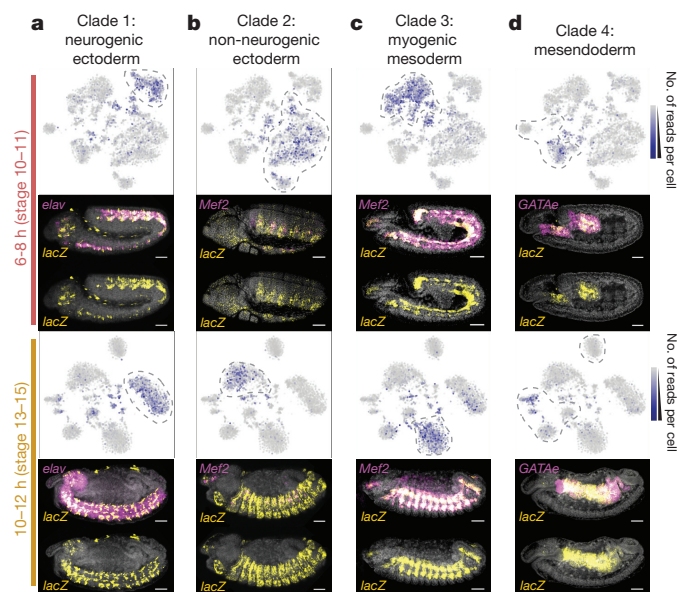


Figure 4 | Prediction of tissue-specific enhancer activity using sci-ATAC-seq. **a–d**, Examples of candidate LSI clade-specific enhancers tested with transgenic reporters. For each time point, upper panels show the *t*-SNE map with blue intensity representing the number of sci-ATAC-seq reads obtained from each tested element. Cell clusters bounded by dashed lines correspond to the predicted clade of activity. Lower panels show transgenic embryos with DAPI-stained nuclei (grey), *in situ* hybridization of the *lacZ* reporter gene driven by the enhancer (yellow), and a tissue marker (magenta). All embryo images are lateral views, with anterior left and dorsal up, and are representative of observations across hundreds of embryos. Scale bar, 50 μ m. The activity and an overview of all tested enhancers are shown in Extended Data Fig. 10.

and spatial information^{29,30} at single-cell resolution has the potential to unlock how an organism's genome encodes its development.

Online Content Methods, along with any additional Extended Data display items and Source Data, are available in the online version of the paper; references unique to these sections appear only in the online paper.

Received 24 June 2017; accepted 7 February 2018.

Published online 14 March 2018.

- Cusanovich, D. A. *et al.* Multiplex single cell profiling of chromatin accessibility by combinatorial cellular indexing. *Science* **348**, 910–914 (2015).
- Maduro, M. F., Meneghini, M. D., Bowerman, B., Broitman-Maduro, G. & Rothman, J. H. Restriction of mesoderm to a single blastomere by the combined action of SKN-1 and a GSK-3 β homolog is mediated by MED-1 and -2 in *C. elegans*. *Mol. Cell* **7**, 475–485 (2001).
- Sethi, A. J., Wikramanayake, R. M., Angerer, R. C., Range, R. C. & Angerer, L. M. Sequential signaling crosstalk regulates endomesoderm segregation in sea urchin embryos. *Science* **335**, 590–593 (2012).
- Rodaway, A. & Patient, R. Mesoderm. An ancient germ layer? *Cell* **105**, 169–172 (2001).
- Thomas, S. *et al.* Dynamic reprogramming of chromatin accessibility during *Drosophila* embryo development. *Genome Biol.* **12**, R43 (2011).
- Bonn, S. *et al.* Tissue-specific analysis of chromatin state identifies temporal signatures of enhancer activity during embryonic development. *Nat. Genet.* **44**, 148–156 (2012).
- Kvon, E. Z. *et al.* Genome-scale functional characterization of *Drosophila* developmental enhancers *in vivo*. *Nature* **512**, 91–95 (2014).
- Gallo, S. M. *et al.* REDfly v3.0: toward a comprehensive database of transcriptional regulatory elements in *Drosophila*. *Nucleic Acids Res.* **39**, D118–D123 (2011).
- Frise, E., Hammonds, A. S. & Celniker, S. E. Systematic image-driven analysis of the spatial *Drosophila* embryonic expression landscape. *Mol. Syst. Biol.* **6**, 345 (2010).
- Tomancak, P. *et al.* Systematic determination of patterns of gene expression during *Drosophila* embryogenesis. *Genome Biol.* **3**, research0088.1 (2002).
- Bonn, S. *et al.* Cell type-specific chromatin immunoprecipitation from multicellular complex samples using BITS-ChIP. *Nat. Protoc.* **7**, 978–994 (2012).

- Doe, C. Q. Temporal patterning in the *Drosophila* CNS. *Annu. Rev. Cell Dev. Biol.* **33**, 219–240 (2017).
- Ciglar, L. & Furlong, E. E. Conservation and divergence in developmental networks: a view from *Drosophila* myogenesis. *Curr. Opin. Cell Biol.* **21**, 754–760 (2009).
- Spahn, P. *et al.* Multiple regulatory safeguards confine the expression of the GATA factor *serpent* to the hemocyte primordium within the *Drosophila* mesoderm. *Dev. Biol.* **386**, 272–279 (2014).
- Reuter, R. The gene *serpent* has homeotic properties and specifies endoderm versus ectoderm within the *Drosophila* gut. *Development* **120**, 1123–1135 (1994).
- Cannavò, E. *et al.* Genetic variants regulating expression levels and isoform diversity during embryogenesis. *Nature* **541**, 402–406 (2017).
- Van Der Maaten, L. & Hinton, G. H. Visualizing data using t-SNE. *J. Mach. Learn. Res.* **9**, 2579–2605 (2008).
- Rodriguez, A. & Laio, A. Machine learning. Clustering by fast search and find of density peaks. *Science* **344**, 1492–1496 (2014).
- Qiu, X. *et al.* Reversed graph embedding resolves complex single-cell developmental trajectories. *Nat. Methods* **14**, 979–982 (2017).
- Lecuit, T., Samanta, R. & Wieschaus, E. *slam* encodes a developmental regulator of polarized membrane growth during cleavage of the *Drosophila* embryo. *Dev. Cell* **2**, 425–436 (2002).
- Beiman, M., Shilo, B. Z. & Volk, T. Heartless, a *Drosophila* FGF receptor homolog, is essential for cell migration and establishment of several mesodermal lineages. *Genes Dev.* **10**, 2993–3002 (1996).
- Okumura, T., Matsumoto, A., Tanimura, T. & Murakami, R. An endoderm-specific GATA factor gene, dGATAe, is required for the terminal differentiation of the *Drosophila* endoderm. *Dev. Biol.* **278**, 576–586 (2005).
- Clark, H. F. *et al.* *Dachsous* encodes a member of the cadherin superfamily that controls imaginal disc morphogenesis in *Drosophila*. *Genes Dev.* **9**, 1530–1542 (1995).
- Simcox, A. A. & Sang, J. H. When does determination occur in *Drosophila* embryos? *Dev. Biol.* **97**, 212–221 (1983).
- Tingvall, T. O., Roos, E. & Engström, Y. The GATA factor *serpent* is required for the onset of the humoral immune response in *Drosophila* embryos. *Proc. Natl Acad. Sci. USA* **98**, 3884–3888 (2001).
- Cao, J. *et al.* Comprehensive single-cell transcriptional profiling of a multicellular organism. *Science* **357**, 661–667 (2017).
- McKenna, A. *et al.* Whole-organism lineage tracing by combinatorial and cumulative genome editing. *Science* **353**, aaf7907 (2016).
- Raj, B. *et al.* Simultaneous single-cell profiling of lineages and cell types in the vertebrate brain by scGESTALT. Preprint at <https://doi.org/10.1101/205534> (2017).
- Karaiskos, N. *et al.* The *Drosophila* embryo at single-cell transcriptome resolution. *Science* **358**, 194–199 (2017).
- Frieda, K. L. *et al.* Synthetic recording and *in situ* readout of lineage information in single cells. *Nature* **541**, 107–111 (2017).
- Tomancak, P. *et al.* Global analysis of patterns of gene expression during *Drosophila* embryogenesis. *Genome Biol.* **8**, R145 (2007).
- Hammonds, A. S. *et al.* Spatial expression of transcription factors in *Drosophila* embryonic organ development. *Genome Biol.* **14**, R140 (2013).

Supplementary Information is available in the online version of the paper.

Acknowledgements This work was technically supported by the EMBL Advanced Light Microscopy, Genomics and Flow Cytometry Facilities. We thank D. Prunkard and L. Gitari in the UW-Pathology Flow Cytometry Facility for their assistance with sorting, and all members of the Furlong and Shendure laboratories for discussions and comments. This work was financially supported by BMBF (TransDiag-2) funds to E.E.M.F., and NIH (DP1HG007811 and R01HG006283) and the Paul G. Allen Family Foundation funds to J.S. D.A.C. was partly supported by T32HL007828 from the National Heart, Lung, and Blood Institute. J.S. is a Howard Hughes Medical Institute Investigator.

Author Contributions D.A.C., J.P.R., D.A.G., J.S. and E.E.M.F. designed the study, explored results and prepared the manuscript, with contributions from all authors. D.A.C. and R.M.D. developed and optimized sci-ATAC-seq, with assistance from L.C. and F.J.S. J.P.R. and D.A.G. led sample preparation and biological validations, with assistance from R.M.-F. D.A.C., J.P.R. and D.A.G. led data analysis, with assistance on specific analyses from D.A., H.A.P., C.T. and X.Q. J.S. and E.E.M.F. supervised the study.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare competing financial interests: details are available in the online version of the paper. Readers are welcome to comment on the online version of the paper. Publisher's note: Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations. Correspondence and requests for materials should be addressed to J.S. (shendure@uw.edu) and E.E.M.F. (furlong@embl.de).

Reviewer Information Nature thanks M. Bulyk, S. Gisselbrecht and B. Gottgens for their contribution to the peer review of this work.

METHODS

Fixation of embryos and nuclear isolation. Wild-type *D. melanogaster* embryos were collected and fixed as previously described³³. In brief, embryos were collected on apple-agar plates in two-hour windows following three one-hour pre-collections to synchronize the collections. After ageing (at 25 °C) to the appropriate time window, embryos were washed from the plates, cleaned and dechorionated in 50% bleach for 2 min, followed by 15-min fixation with shaking at room temperature in cross-linking solution (50 mM Hepes, 1 mM EDTA, 0.5 mM EGTA, 100 mM NaCl, pH 8, 1.8% formaldehyde v/v) with a heptane layer. Fixation was stopped by washing with 125 mM glycine in PBS. The embryos were washed, dried and frozen at -80 °C in ~1-g aliquots. Embryo dissociation and nuclear isolation were performed as described previously (steps 1–10)¹¹ using a dounce homogenizer and a 22G needle. The resulting nuclei were pelleted at 2,000g at 4 °C, resuspended in nuclear freezing buffer (50 mM Tris at pH 8.0, 25% glycerol, 5 mM Mg(OAc)₂, 0.1 mM EDTA, 5 mM DTT, 1 × protease inhibitor cocktail (Roche), 1:2,500 superasin (Ambion)) and flash frozen in liquid nitrogen.

Collection of sci-ATAC-seq data. Our protocol for generating sci-ATAC-seq data was largely as previously described¹, but with a few important improvements. Frozen nuclei were thawed quickly in a 37 °C water bath and then pelleted at 500g for 5 min at 4 °C, aspirated and resuspended in cold lysis buffer (supplemented with protease inhibitors). Nuclei were stained with 3 μM DAPI and 2,500 DAPI⁺ nuclei were sorted into each well of a 96-well plate containing 9 μl lysis buffer (10 mM Tris-HCl, pH 7.4, 10 mM NaCl, 3 mM MgCl₂ and 0.1% IGEAL CA-630³⁴, supplemented with protease inhibitors (Sigma)) and 10 μl TD buffer (Illumina, part of FC-121-1031) in each well. One microlitre of each of the 96 custom and uniquely indexed Tn5 Transposomes (Illumina, 2.5 μM)³⁵ was then added to each well and nuclei were incubated at 55 °C for 30 min. Following tagmentation, 20 μl 40 mM EDTA (supplemented with 1 mM spermidine) was added to stop the reaction and the plate was incubated at 37 °C for 15 min. All wells of the plate were then pooled, nuclei were stained again with 3 μM DAPI and 25 DAPI⁺ nuclei were sorted into each well of a second set of 96-well plates that contained 12 μl reverse crosslinking buffer (11 μl EB buffer (Qiagen) supplemented with 0.5 μl 20 mg/ml Proteinase K (Qiagen) and 0.5 μl 1% SDS). For each time point, we collected four plates of nuclei at this stage. We expect that sorting 25 nuclei into each well at this stage will result in approximately 12% of barcodes representing more than one nucleus (collisions)¹. Nuclei were then incubated overnight at 65 °C. Proceeding from reverse-crosslinking, we added primers (0.5 μM final concentration, Supplementary Table 12), 7.5 μl NPM polymerase master mix (Illumina, FC-121-1012) and BSA (2 × final concentration; NEB) to each well. Tagmented DNA was then PCR amplified. To determine the number of cycles required, we first amplified several test wells of nuclei that had been sorted onto an additional plate and monitored the reactions with SYBR green on a qPCR machine to establish when the libraries reached saturation. The cycling conditions were as follows: 72 °C 3 min, 98 °C 30 s; 98 °C 10 s, 63 °C 30 s, 15–25 cycles; 72 °C 1 min, hold at 10 °C.

We have found that the optimal number of cycles can vary from one experiment to the next, but is usually in the range of 15–25 cycles. After PCR amplification, all wells were pooled and split across four DNA Clean & Concentrator-5 columns (Zymo) and all four products were then pooled and cleaned again using Ampure beads (Agencourt). Finally, the concentration and quality of the libraries was determined using the BioAnalyzer 7500 DNA kit (Agilent). For sequencing, equimolar libraries from the three time points were pooled and loaded at 1.5 pM on a NextSeq High output 300 cycle kit and sequenced using custom primers and a custom sequencing recipe³⁵. Fifty base pairs were sequenced from each end, in addition to the barcodes introduced during tagmentation and PCR amplification. This improved protocol resulted in roughly an order of magnitude more reads per cell than previously reported.

Read alignment, cell assignment and duplicate removal. To process the data, BCL files were converted to fastq files using bcl2fastq v.2.16 (Illumina). Each read was assigned a barcode which was actually made up of four individual components: a tagmentation barcode and a PCR barcode added to the P5 end of the molecule, and a distinct tagmentation and PCR barcode added to the P7 end of the molecule. To correct for sequencing and/or PCR amplification errors, we broke the barcode into its constituent parts and matched each piece against all possible barcodes. If the component was within three edits of an expected barcode and the next best matching barcode was at least two edits further away, we fixed the barcode to its presumptive match. Otherwise, we classified the barcode as ambiguous or unknown. We next mapped each read to the dm3 reference genome using bowtie2³⁶ with '-X 2000 -3 1' as options and filtered out read pairs that did not map uniquely to autosomes or sex chromosomes with a mapping quality of at least 10, as well as reads that were associated with ambiguous or unknown barcodes. Of 430,658,635 sequenced read pairs, 301,314,040 (70%) mapped to the nuclear reference genome with an assigned cell barcode. By contrast, only 366,468 read pairs (0.09%) mapped to the mitochondrial genome, with an assigned cell

barcode. We subsequently removed PCR duplicates for all reads that mapped to the nuclear genome using a custom Python script that only considered reads assigned to the same barcode. Finally, to determine which barcodes represented genuine cells (as opposed to background reads assigned to improper barcodes), we counted the number of reads assigned to each barcode and log-transformed those counts and then used the mclust package in R^{37,38}, which fits the data using a mixture model and determines the maximum likelihood parameters for a given number of distributions, to define two distributions of barcodes—setting the read depth cut-off for a cell at the point at which we were 95% confident that the barcode belonged to the higher read-depth distribution. Considering the distribution of barcodes for all three experiments at the same time, we determined this read-depth cut-off to be 500 reads (that is, we required a barcode to be associated with at least 500 reads to be considered a true cell; Extended Data Fig. 1). See <http://atlas.gs.washington.edu> for more details on data processing.

Latent semantic indexing. To further process the raw data, we first broke the genome into 2-kb windows and then scored each cell according to whether it had any insertions in each window, creating a large binary matrix of windows by cells for each time point. Based on this binary matrix, we retained only the top 20,000 most commonly used sites (this number could extend a little above 20,000 because we retained all sites that were tied at the threshold for cell counts) and then filtered out the 10% of cells with the smallest number of accessible sites. We then normalized and re-scaled these large binary matrices by using the term frequency-inverse document frequency (TF-IDF) transformation. We first weighted each site that was accessible in an individual cell by the total number of sites accessible in that cell. We then multiplied these weighted values by log(1 + the inverse frequency of each site across all cells). Subsequently, we performed singular value decomposition on the TF-IDF matrix and then generated a lower-dimensional representation of the data by only considering the second to sixth dimensions (because we have found that the first dimension is always highly correlated with read depth). These LSI scores were then used to cluster cells and windows on the basis of cosine distances using the ward algorithm in R. Scores of accessibility were standardized by row and capped at ± 1.5 for visualization. Visual examination of the resulting bi-clustered heat map identified 4–5 major clades for each time point.

Peak calling. To identify specific regulatory elements within each of the major clades at each time point, we aggregated the data across cells from each clade using a process we call 'in silico cell sorting'. To do so we collected all the unique mapped reads associated with cells that were assigned to a given clade and saved them as a distinct bam file. Then for each bam file representing a clade, we used MACS2³⁹ to identify peaks of increased insertion frequency, as well as summits of accessibility within each of those peaks. For MACS, we used the macs2 callpeak command with the following parameters: "--nomodel --keep-dup all --extsize 200 --shift -100 --format BAM --gsize mm --call-summits". For downstream analyses we generated a master list of potential regulatory elements by taking 150-bp windows centred on all summits called in each clade in each time point and merged them with the BEDTools program⁴⁰. For Extended Data Fig. 1d, we also compared our sci-ATAC-seq data to previously collected DNase-seq bulk data⁵ on whole embryos at similar time points. To be consistent in our comparisons (and provide a comprehensive list of peaks), we downloaded the raw DNase-seq reads (36 bp, single-end), remapped them with our pipeline and called peaks with MACS2 as described above. Specifically, we downloaded two replicates for each of three time points: stage 5, stage 11 and stage 14. Peaks called on each replicate independently were intersected to create a master list of peaks for each time point, which were then intersected with our sci-ATAC-seq data.

Identification of differentially accessible sites. To identify regulatory elements that were more specifically accessible in individual clades, we generated a new binary matrix of insertion scores for individual cells using the master list of summits of accessibility described above. We then used a logistic regression framework to test whether cells of a given clade were more likely to have insertions at a given site relative to all other cells. To identify sites that were specifically more accessible in a single clade, we first found summits that were significantly more open in a given clade at a 1% FDR, including log₁₀(total unique reads) for each individual cell as a covariate. To ensure that these sites were specific to any one clade, we also filtered out sites that were significantly accessible in any other clade at a relaxed 20% FDR. All testing of differential accessibility was implemented with the Monocle 2 package^{19,41} using the binomial test. For this analysis, only sites observed in at least 50 cells in a given time point were tested.

k-mer discovery. We used SeqGL⁴² to identify motifs that were enriched in clade-specific elements. To do so, we started with all clade-specific sites, based on our logistic regression testing described above. Because our master list of sites included sites of variable length (after merging all sites from all clusters), we only considered 150-bp windows centred on summit midpoints. We also removed sites within 500 bp of a transcription start site (TSS), to focus on tissue-specific distal elements. As a background set of regions we randomly selected an equal number of

sites from the master summit list that matched the GC and repeat element content of the test set (this was controlled using a script provided in the gkm-SVM software package)⁴³. Finally, instead of default parameters, we used 200 groups and 30,000 features, similar to the parameters used to analyse DNase-seq data in the original SeqGL publication⁴².

Enrichments for tissue or cell-type activity and transcription factor binding data. To perform categorical enrichments, we annotated regions, windows and peaks of the non-coding genome using two types of experimental information: (1) tissue-specific expression of the nearest gene comprising *in situ* hybridization data from the Berkeley *Drosophila* Genome Project (<http://insitu.fruitfly.org/cgi-bin/ex/insitu.pl>) and a download of the FlyBase gene-expression annotations (May 2016); (2) a custom enhancer database of ~8,000 transgenic reporter assays covering 15% of the non-coding genome, containing spatio-temporal information of ~4,000 active developmental enhancers (CAD4; Supplementary Table 13). We compiled the enhancer database (CAD4) from three primary resources: our previous CRM Activity Database (CAD)⁶, entries from the RedFly enhancer database (Release 5)⁸, and data from the Vienna Tiling Project⁷. We compiled this dataset in two steps. First, all expression terms (and timing terms, where available) were mapped to a common standard (FlyBase anatomy terms v.1.47) and, when timing information was available, a common set of stage windows (stages 1–3, stages 4–6, stages 7–8, stages 9–10, stages 11–12, stages 13–16). In most cases, the mapping was automatic and unambiguous. In some cases, manual term matching was required (generally unambiguous). In the second step, we merged overlapping entries from CAD3 and the RedFly database and manually removed redundant information. Given the different methodologies used in the compilation of the data sources, no attempt was made to merge entries from CAD3/RedFly with the Vienna Tiles.

Almost all expression terms for both the gene and enhancer annotations could be mapped to a common set of hierarchically organized anatomical terms (FlyBase anatomy OBO file v.1.47). In the few cases where an exact match could not be found, a choice was made manually or using the map provided by FlyBase (FBrf0219073). The stage or timing information from both datasets was shifted as needed to match a common set of grouped stages (stages 1–3, stages 4–6, stages 7–8, stages 9–10, stages 11–12, stages 13–16). The compiled data are shown in Supplementary Table 13. In addition to BDGP/FlyBase gene expression data, we made use of *Drosophila*-specific gene-level functional information (biological process, molecular function and cellular compartment) downloaded from the Gene Ontology Consortium (v.1.2) and additional, higher-level functional annotations downloaded from the PANTHER classification system (v.8) corresponding roughly to the higher-level categories of the GO-SLIM ontology.

To further explore the functions of specific regions of noncoding DNA, we also made use of a custom compilation of high-quality transcription factor binding data from ChIP studies during embryogenesis (taken from ref. 16) that allowed us to assign transcription factor binding events to each sciATAC window or peak. Transcription factor binding motifs were taken from this same dataset. To infer likely transcription factor binding events, we scanned under published ChIP peaks for instances of the motif using FIMO⁴⁴. Enrichments for these data are listed under the category name 'custom' in the enrichment data tables.

Categorical enrichments. To identify enriched categories within the LSI clades, we first assigned categorical labels by looking for overlaps between our summit regions and our enhancer activity database, with summits inheriting the timing and expression labels of all overlapping enhancers. Gene-based annotations (expression, GO and PANTHER terms) were assigned by association to the nearest gene.

To identify differentially accessible summit regions, we used a logistic-regression framework (see above) as applied to all summit regions containing reads in at least 50 cells. Enriched summit regions constituted the foreground set for any clade, with the remaining tested summit regions constituting the background set. For each of our category sets (for example, enhancer expression, gene expression or GO), we used a Fisher's exact test to look for over-representation of each category among our foreground set relative to the background set. Because many of our categories are strongly overlapping, we have applied no formal correction for multiple comparison, choosing instead to focus on large, consistent enrichments with highly significant *P* values. Overlaps among significant categories were visualized by plotting distances between categories using the pyEnrichment package (<https://github.com/ofedrigo/pyEnrichment>) to avoid overcalling a category.

Categorical enrichment within our *t*-SNE clusters was assessed similarly. Foreground sets per cluster (within each time point) were assessed using the results of our binomial enrichment test (*q* value ≤ 0.01 and a $\beta > 0$). The background set consisted of all other tested summits at that time point (see above).

***t*-SNE and cluster identification.** To identify clusters of cells with finer resolution than the LSI-based clades, we used *t*-SNE¹⁷ for dimensionality reduction. We started with the same binary matrix of insertions in summits that we used to identify clade-specific differentially accessible sites. We again filtered out the lowest

10% of cells (in terms of site coverage) and in this case we retained only sites that were observed in at least 5% of cells. We then transformed this matrix with the TF-IDF algorithm described above. Finally, we generated a lower-dimensional representation of the data by including the first 50 dimensions of the singular value decomposition of this TF-IDF-transformed matrix. This representation was then used as input for the Rtsne package in R^{17,45,46}. To identify clusters of cells in this 2D representation of the data, we used the density peak clustering algorithm¹⁸ as implemented in Monocle 2^{19,41}. Rho and delta parameters were chosen to be very inclusive of outlier peak centres (based on the decision plot), while making sure that the clusters were sensible based on visual inspection of the cluster assignments on the *t*-SNE plot.

***t*-SNE differential accessibility.** To identify summits that were significantly more accessible in *t*-SNE-defined cell clusters, we used a similar framework to the one described for LSI-based clades above. There were, however, a few differences. In this case, we consider sites that were seen in at least 10 cells in any time point (instead of 50). In addition, we did not use a second cut-off to determine specificity within a time point.

Sexing individual nuclei. Another biological axis of the data that came to light through the use of *t*-SNE plots was that we were able to clearly distinguish nuclei from male and female embryos. In an initial analysis, we included data from the sex chromosomes while clustering cells (as was done for the germ-layer analysis). This resulted in many individual cell clusters appearing 'bi-lobed' (Extended Data Fig. 5a), which prompted us to explore whether there was sex bias in the lobes of individual cell clusters. We found that the distribution of reads mapping to the X chromosome in individual cells was distinctly bimodal (Extended Data Fig. 5b), allowing us to assign a sex to each cell. When we coloured the *t*-SNE plots according to these sex assignments we found that the lobes of individual cell clusters almost perfectly segregated the sexes (Extended Data Fig. 5c). Although this may be very useful for future studies, we alleviated this bi-lobed problem here by excluding sex chromosome reads from our analysis and re-clustered cells with *t*-SNE. This resolved the bi-lobed structures and removed the sex bias from almost every individual cluster (Extended Data Fig. 5d).

Arranging single cells from 2–4-h embryos along developmental trajectories. Because we noted that cells from 2–4-h embryos were distributed across the *t*-SNE map in a manner consistent with their developmental stage, we sought to more formally evaluate the arrangement of individual cells along a temporal trajectory. We used Monocle 2^{19,41} v.2.5.3, which uses a reverse graph embedding algorithm to learn trajectories in single-cell data and was recently extended to single-cell ATAC-seq data⁴⁷. To define sites to use for ordering cells, we combined the *t*-SNE clusters into major groups on the basis of our annotations—blastoderm, mesoderm, endoderm, ectoderm, neural ectoderm, unknown and collisions—and identified sites that were differentially accessible (1% FDR) between each cluster and all other cells within that time point (with the exception of the collision and unknown clusters). We then took the union of sites that were among the 100 most differentially accessible for each cluster and used this set of sites to order cells in Monocle. In order to reduce the sparsity of the data, we aggregated all sites that were within 1 kb of each other and summed their reads to obtain a regional score accessibility. Using these aggregated sites as features, cells were ordered by the DDRTree algorithm in three dimensions ('max_components = 3'), with the ncenter parameter set to 200 and the maxIter parameter set to 1,000 during the dimensionality-reduction step. Only the first two dimensions are visualized and the coordinates of the first dimension were multiplied by -1 so that pseudotime would run from left to right (Fig. 2c). This resulted in a tree with four differentiated branches representing the major germ layers (one is a possibly spurious, short branch along the ectodermal lineage). On the basis of this ordering, we aimed to identify sites that were significantly associated with progression in pseudotime using the likelihood-ratio testing framework in Monocle 2 (Supplementary Table 6). As with ordering the cells, we adopted a strategy to reduce the sparsity of our data. Specifically, we binned the pseudotime into 100 bins and counted how many cells had accessible chromatin in each pseudotime bin for each site. All sites that were accessible in more than ten cells were tested. To identify sites that were associated with pseudotime in a lineage-specific fashion we used a similar framework. First, we separated out cells along each unbranched path through the trajectory to test separately for pseudotime dependence. We took the cells at the tip of each lineage state and traversed the graph to the root state (that is, beginning of the pseudotime), collecting the cells that were arranged along this path. As mentioned above, there was a small branch off of the ectodermal lineage that was ignored for this analysis. Then we binned the cells along this single pseudotime branch and performed likelihood ratio testing for each lineage as we did for the global pseudotime measure (Supplementary Tables 7–9). After testing all three lineages, we defined a site as specific to a lineage if it was significantly associated with pseudotime in that lineage (1% FDR) but was not significantly associated with pseudotime in the other two lineages at a relaxed threshold (20% FDR).

Identifying clusters of cells that are likely artefacts of barcode collisions. Several small clusters (for example, cluster 6 at 6–8 h) appear to be mixtures of cells from different germ layers and/or tissues, based on our enrichment analysis. To determine whether these were technical (due to barcode collisions, where one cell barcode represents the nuclear contents of two cells) or biological, we used two metrics to identify collisions (instances wherein two or more cells coincidentally pass through the same combination of wells during sci-ATAC-seq). First, we looked at the estimated complexity of individual cells that make up these small clusters, as collisions are expected to be twice as complex on average as barcodes that truly represent an individual cell. To calculate the estimated library complexity (that is, the estimated total number of unique reads per cell in the library), we used the same algorithm as implemented in Picard (<http://broadinstitute.github.io/picard>) on a cell-by-cell basis. Second, we considered whether the proportion of reads mapping to the X chromosome for cells in these clusters was distinctly bimodal, as collisions would be just as likely to combine data from cells of the opposite sex as from two cells of the same sex (Extended Data Fig. 7). While the vast majority of clusters exhibited distributions of complexity and X chromosome coverage consistent with single nuclei, a small subset of clusters in each time point showed either higher complexity than expected, more unimodality of reads mapping to the X chromosome, or both—consistent with our suspicion that these are cell collision clusters (Extended Data Fig. 7). At 2–4 h, we identified one (2.3% of cells), at 6–8 h we identified three (5.8% of cells) and at 10–12 h we identified six (7.3% of cells) potential collision clusters (Figs 2a, 4a, b, purple clusters).

Transgenic enhancer assays. Candidate clade-specific enhancers were selected from sci-ATAC-seq summits using the following criteria only: (1) summit shows enriched accessibility specifically in the target cell clade at 6–8 h and/or 10–12 h (q value < 0.01 and $\beta > 0$ in target clade, q value > 0.2 in all other clades); (2) summit does not fall within 500 bp of an annotated transcription start site; (3) summit does not overlap a region already in our database of characterized developmental enhancers. Summits showing a range of effect sizes (β) were selected (minimum β approximately 1.9; see Supplementary Table 11). The selected regions, plus 100–200 bp of flanking sequence, were PCR amplified from genomic DNA (primers are listed in Supplementary Table 11) and cloned upstream of a minimal *hsp70* promoter driving a *LacZ* reporter gene in an attB-containing plasmid. All constructs were injected into embryos according to standard methods⁴⁸ and inserted into the attP landing site line M{3×P3-RFP.attP}ZH-51C via PhiC31 integrase insertion⁴⁹, yielding integration at chromosomal position 51C1. Transgenic lines were generated by BestGene. Ten elements from each of the four germ-layer clades were initially selected—some failed at the cloning or transgenesis phase. We obtained 31 transgenic lines, representing six candidate regions with specific accessibility in neurogenic ectoderm, ten in non-neurogenic ectoderm, eight in myogenic mesoderm and seven in non-myogenic mesoderm plus endoderm.

Overnight collections of homozygous embryos spanning all stages of embryogenesis were formaldehyde-fixed, stained by double fluorescent *in situ* hybridization⁵⁰, and mounted in ProLong Gold with DAPI (Invitrogen; cat. #P36931). Antisense *in situ* probes against *LacZ* and a tissue marker gene were used: *Mef2*-marking myogenic mesoderm was used for predicted myogenic mesoderm and non-neurogenic ectoderm enhancers; *GATAe* was used for predicted non-myogenic mesoderm and endoderm enhancers. For the predicted neurogenic ectoderm enhancers, neurons were marked by immunostaining with antibodies against the Elav protein (Elav-9F8A9; Developmental Studies Hybridoma Bank). The annotation of enhancer activity is based on observations across hundreds of embryos. Representative images were acquired with a Zeiss LSM780 laser-scanning confocal microscope using a PlanApo 20×/NA 0.8 objective at an effective pixel size of 461 nm in the x - y plane. Images were processed using Fiji⁵¹. Annotated t -SNE plots for each candidate enhancer were produced by plotting the sum of sci-ATAC-seq reads per cell that overlapped each tested genomic region.

FACS isolation of tissue-specific nuclei and DNase-seq. Target populations of cell nuclei from staged fixed embryos were obtained by FACS as previously described¹¹ with the following modifications. Prior to incubation with primary antibodies, nuclei from 6–8 h embryos were incubated in PBS supplemented with 5% BSA, 0.1% TritonX-100 and 0.2% Igepal-630 on a rotator at 4°C for 30 min. Primary antibody staining was performed overnight at 4°C in 3 ml PBS supplemented with 5% BSA and 0.1% TritonX-100 per 1 g frozen embryos. Primary antibodies used were monoclonal anti-Elav (Developmental Studies Hybridoma Bank 9F8A9 at 1:100 dilution) to mark postmitotic neurons and anti-Mef2 (produced and pre-cleared in the Furlong laboratory and used at 1:200 dilution) to mark myogenic mesoderm. Secondary antibody staining was performed for 1 h at 4°C in the same buffer. Following each antibody staining, nuclei were washed twice by pelleting and resuspending in 10 ml PBS supplemented with 5% BSA. An aliquot of stained, unsorted nuclei was put aside to represent the whole embryo. For DNase digestion, nuclei were resuspended in R buffer (7.5 mM Tris pH8, 45 mM NaCl, 30 mM

KCl, 6 mM MgCl₂, 1 mM CaCl₂) and 10–20 million nuclei were digested using 5–20 U DNaseI at 37°C for 3 min, and the reaction was stopped by adding 500 μ l stop buffer (50 mM Tris pH8, 100 mM NaCl, 0.1% SDS, 100 mM EDTA pH8). A small control digest without DNaseI was performed to assess DNA integrity. Following addition of RNaseA, samples were incubated at 55°C for 10 min, then 25 μ l proteinase K (25 mg/ml) was added and the samples were incubated overnight at 65°C to reverse cross-links. A small aliquot was run on a 1% agarose gel to assess digestion levels, and optimal digests were size-fractionated using 10–40% sucrose gradients. DNA fragments ~100–500 bp in length were isolated from fractions using a Qiagen PCR clean up kit and checked for enrichment in known hypersensitive sites by qPCR. The digests with the highest qPCR enrichment were selected for library preparation using the NextFlex qRNA-seq Kit v.2 (Biooscientific #NOVA-5130-12). In brief, ~10–30 ng DNA consisting of ~100–500 bp fragments that result from DNase digestion was end-repaired and terminal adenosine residues were added. Adapters containing in-line molecular barcodes were ligated, after which the material was size selected using AMPure beads (negative selection with 0.6× beads, then positive selection with 0.98× beads). PCR amplification was performed using barcoded primers to introduce sample barcodes for 12–16 cycles, depending on input amount. The PCR-amplified library was purified using AMPure beads, quantified using a Qubit High-sensitivity DNA kit (Invitrogen), and sized on a Bioanalyzer High-Sensitivity DNA chip (Agilent). Libraries were pooled and sequenced in paired-end mode on a HiSeq2000 (Illumina). Reads were mapped to the Dm3 reference genome using BWA aln⁵², keeping only reads with a mapping quality score greater than 20. Duplicate reads originating from PCR were removed using the Je suite⁵³ making use of the molecular indices.

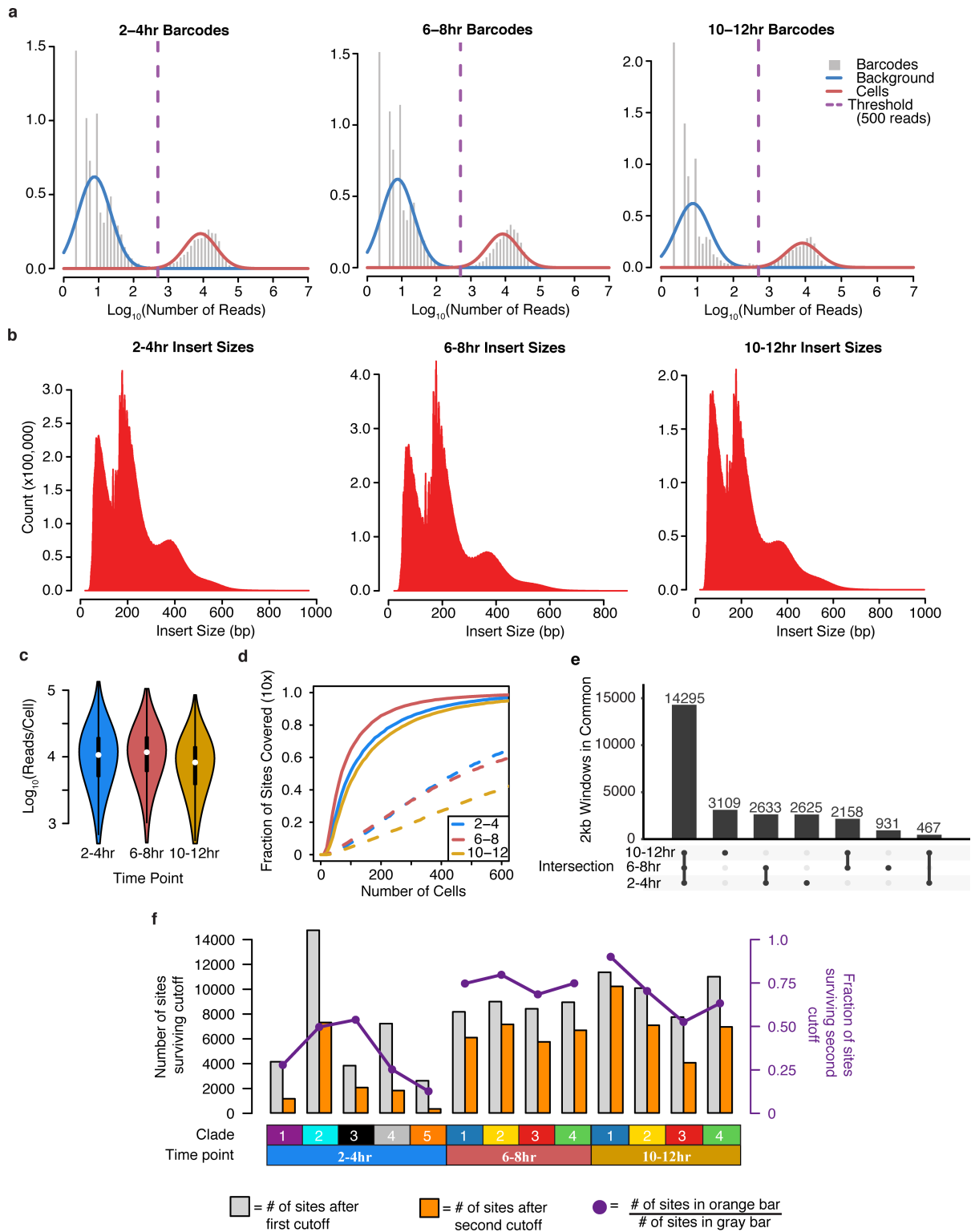
Ethics statement. Anti-Mef2 antibodies were generated from rabbits at EMBL in accordance with European Law and EMBL ethical guidelines. *Drosophila melanogaster* were reared and collected at EMBL in accordance with standard practice and the ethical standards of the European research community.

Code availability. Most of the code used in processing and analysis of the data in this article is available at <http://atlas.gs.washington.edu>. Any code not provided there will be made available upon request.

Data availability. All raw ATAC-seq and DNase-seq data are available through GEO (accession GSE101581) and ArrayExpress (E-MTAB-5999). BigWig files for coverage within each clade, regions of accessibility (peak calls) and a master list of all potential regulatory elements (Supplementary Table 1) will be made available on the Furlong laboratory web page (<http://furlonglab.embl.de/data>). To make the data easily accessible we have generated a searchable html page where users can select a t -SNE cluster or genomic locus of interest and visualize the data throughout the genome (<http://shiny.furlonglab.embl.de/scATACseqBrowser/>) and site-by-cell matrices and vignettes to facilitate further exploration of the data (<http://atlas.gs.washington.edu>).

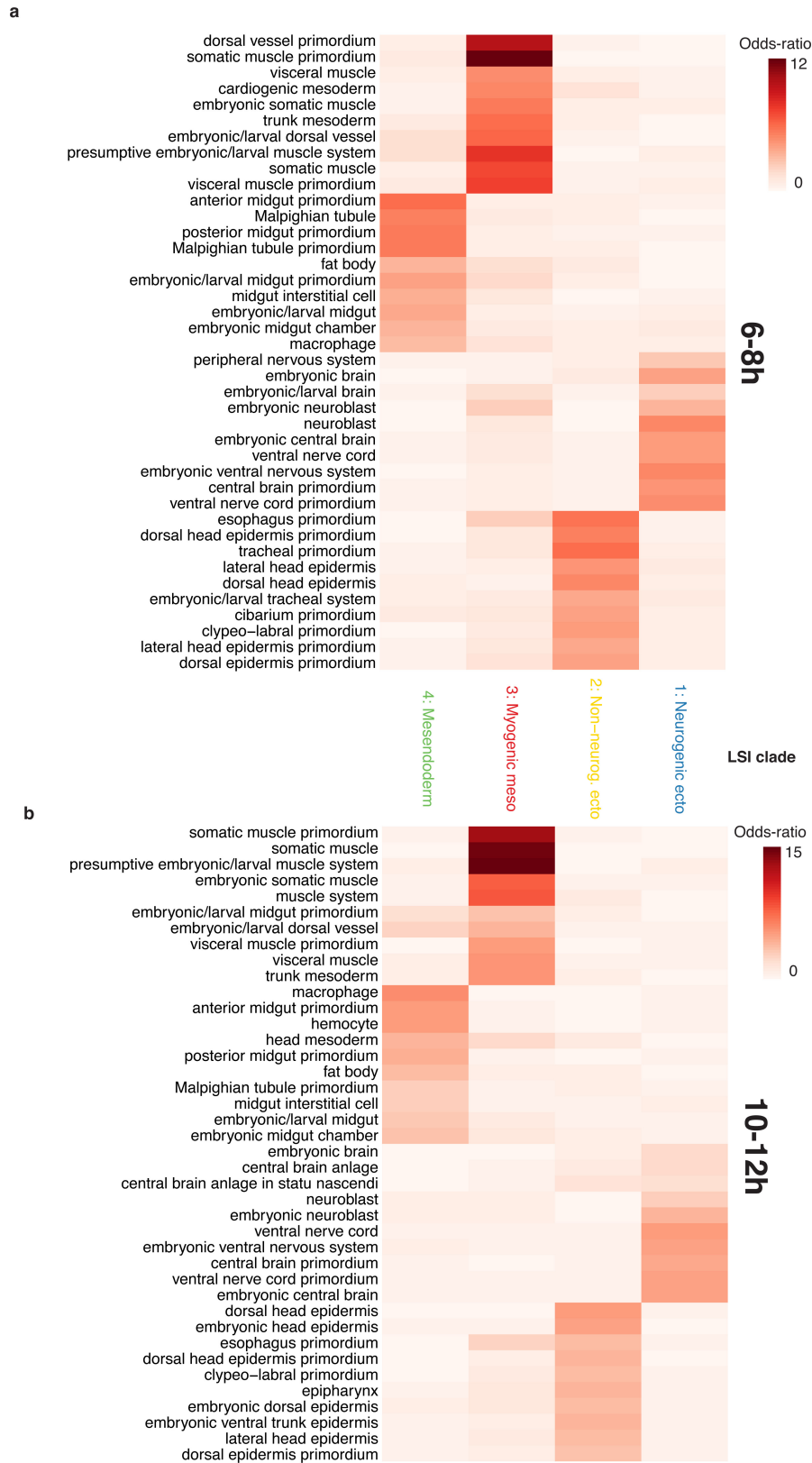
33. Sandmann, T., Jakobsen, J. S. & Furlong, E. E. ChIP-on-chip protocol for genome-wide analysis of transcription factor binding in *Drosophila melanogaster* embryos. *Nat. Protoc.* **1**, 2839–2855 (2006).
34. Buenrostro, J. D., Giresi, P. G., Zaba, L. C., Chang, H. Y. & Greenleaf, W. J. Transposition of native chromatin for fast and sensitive epigenomic profiling of open chromatin, DNA-binding proteins and nucleosome position. *Nat. Methods* **10**, 1213–1218 (2013).
35. Amini, S. *et al.* Haplotype-resolved whole-genome sequencing by contiguity-preserving transposition and combinatorial indexing. *Nat. Genet.* **46**, 1343–1349 (2014).
36. Langmead, B. & Salzberg, S. L. Fast gapped-read alignment with Bowtie 2. *Nat. Methods* **9**, 357–359 (2012).
37. Fraley, C. & Raftery, A. E. Model-based clustering, discriminant analysis and density estimation. *J. Am. Stat. Assoc.* **97**, 611–631 (2002).
38. Fraley, C., Raftery, A. E., Murphy, T. B. & Scrucca, L. *Version 4 for R: Normal Mixture Modeling for Model-Based Clustering, Classification, and Density Estimation Technical Report No. 597* (Department of Statistics, Univ. of Washington, 2012).
39. Zhang, Y. *et al.* Model-based analysis of ChIP-seq (MACS). *Genome Biol.* **9**, R137 (2008).
40. Quinlan, A. R. & Hall, I. M. BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics* **26**, 841–842 (2010).
41. Trapnell, C. *et al.* The dynamics and regulators of cell fate decisions are revealed by pseudotemporal ordering of single cells. *Nat. Biotechnol.* **32**, 381–386 (2014).
42. Setty, M. & Leslie, C. S. SeqGL identifies context-dependent binding signals in genome-wide regulatory element maps. *PLOS Comput. Biol.* **11**, e1004271 (2015).
43. Ghandi, M., Lee, D., Mohammad-Noori, M. & Beer, M. A. Enhanced regulatory sequence prediction using gapped k -mer features. *PLOS Comput. Biol.* **10**, e1003711 (2014).
44. Grant, C. E., Bailey, T. L. & Noble, W. S. FIMO: scanning for occurrences of a given motif. *Bioinformatics* **27**, 1017–1018 (2011).
45. Van Der Maaten, L. Accelerating t -SNE using tree-based algorithms. *J. Mach. Learn. Res.* **15**, 3221–3245 (2014).

46. Krijthe, J. H. Rtsne: t-distributed stochastic neighbor embedding using a Barnes–Hut implementation. <https://github.com/jkrijthe/Rtsne> (2015).
47. Pliner, H. *et al.* Chromatin accessibility dynamics of myogenesis at single cell resolution. Preprint at <https://doi.org/10.1101/155473> (2017).
48. Rubin, G. M. & Spradling, A. C. Genetic transformation of *Drosophila* with transposable element vectors. *Science* **218**, 348–353 (1982).
49. Bischof, J., Maeda, R. K., Hediger, M., Karch, F. & Basler, K. An optimized transgenesis system for *Drosophila* using germ-line-specific φ C31 integrases. *Proc. Natl Acad. Sci. USA* **104**, 3312–3317 (2007).
50. Furlong, E. E., Andersen, E. C., Null, B., White, K. P. & Scott, M. P. Patterns of gene expression during *Drosophila* mesoderm development. *Science* **293**, 1629–1633 (2001).
51. Schindelin, J. *et al.* Fiji: an open-source platform for biological-image analysis. *Nat. Methods* **9**, 676–682 (2012).
52. Li, H. & Durbin, R. Fast and accurate short read alignment with Burrows–Wheeler transform. *Bioinformatics* **25**, 1754–1760 (2009).
53. Girardot, C., Scholtalbers, J., Sauer, S., Su, S. Y. & Furlong, E. E. Je, a versatile suite to handle multiplexed NGS libraries with unique molecular identifiers. *BMC Bioinformatics* **17**, 419 (2016).



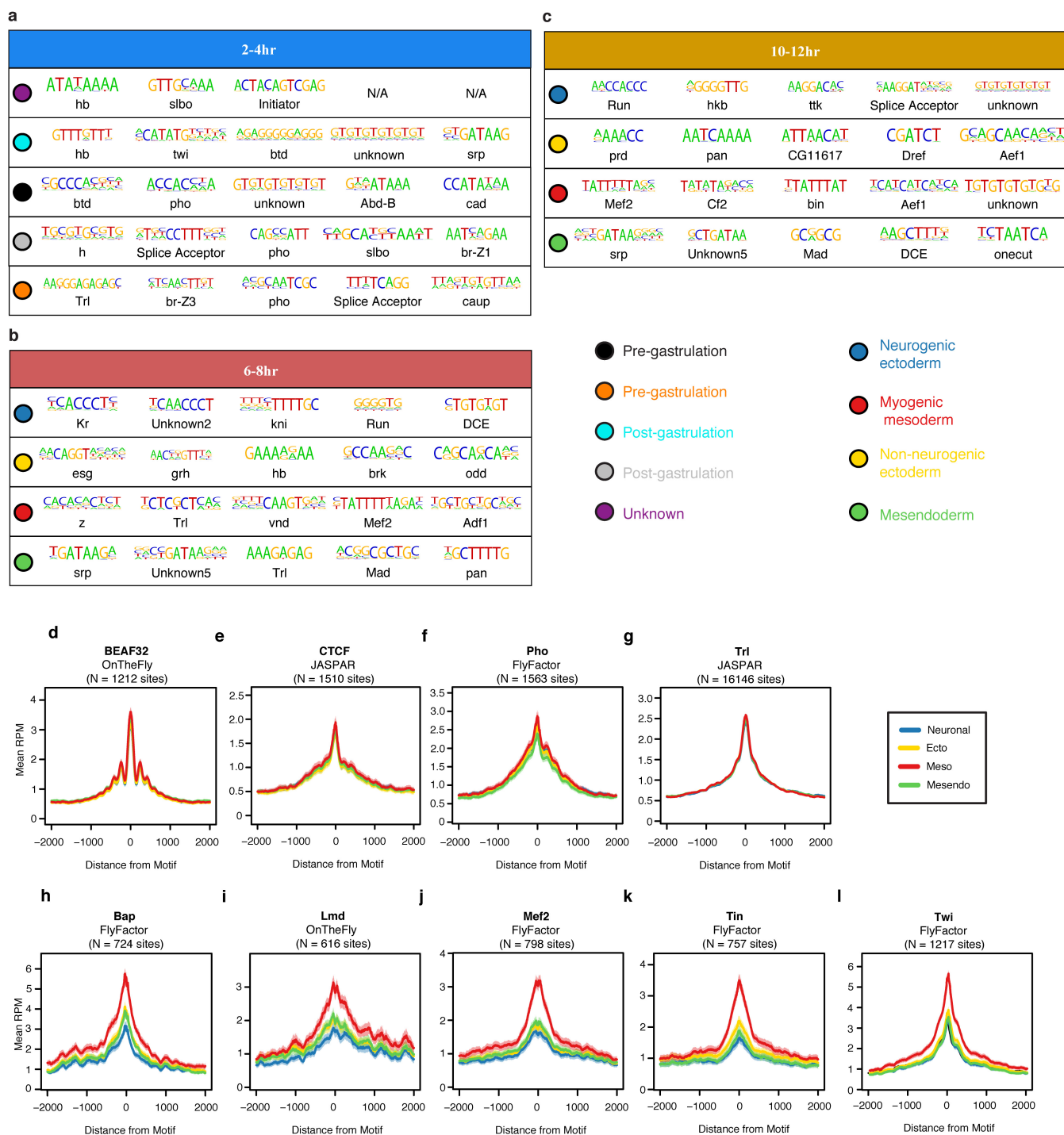
Extended Data Figure 1 | Summary of read distributions across the three sampled time points. a, \log_{10} counts of sci-ATAC-seq reads per barcode at each time point are bimodally distributed. A threshold of 500 reads was used to identify barcodes corresponding to valid cells versus background. **b**, Fragment size distribution at each time point is consistent with expected nucleosomal banding pattern of standard (bulk) ATAC-seq experiments. **c**, Violin plot for distribution of unique, mappable reads per cell at each time point (2–4h, $n = 8,024$; 6–8h, $n = 7,880$; 10–12h, $n = 7,181$) plotted on a logarithmic scale. White point indicates median value, thick black line extends to 25th and 75th percentile, and thin black lines extend to most extreme values within 1.5 times the interquartile range of the median. The filled colour width represents a density estimate

of the distribution of cells along the y axis. **d**, Fraction of previously characterized DHS covered in at least 10 cells upon sampling a given number of cells (solid lines) as compared to random genomic windows (dashed lines). **e**, An UpSet plot shows the degree to which the top 20,000 windows overlap between the three time points. Each bar shows the number of sites included in a specific intersection and the ‘peg board’ below shows which particular comparison is included in that bar. **f**, Bar plot of the number of sites identified as significantly open in each clade (1% FDR; grey bar, first cutoff) and the number of sites specific to that clade (orange bar, second cutoff). Overlaid on the barplot (purple points) is the fraction of sites passing the first cut-off that also pass the second cut-off (count of orange bar/count of grey bar).



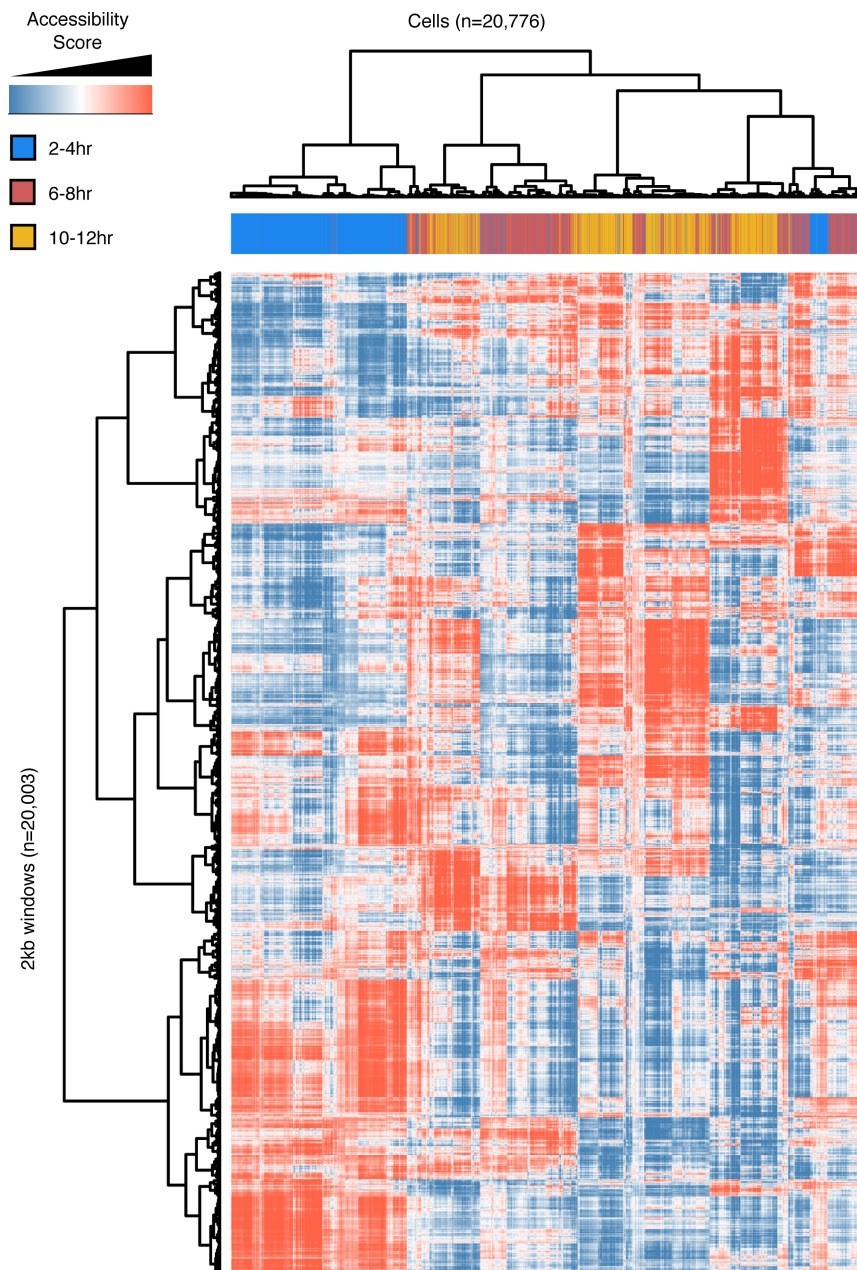
Extended Data Figure 2 | Enhancer enrichments for LSI clades at 6–8 h and 10–12 h. Enrichment for tissue-of-expression information for characterized distal enhancers overlapping clade-specific peaks at 6–8 h (a) and 10–12 h (b). Each column represents a different clade and each row represents an annotation term assigned to tested enhancer elements.

Shading indicates the odds ratio for the intersection of enhancers sharing a given annotation with clade-specific accessible sites. Shown are all categories in the top ten enrichments of any clade (enrichment scores capped at 15 for display) containing at least 35 known enhancer overlaps.



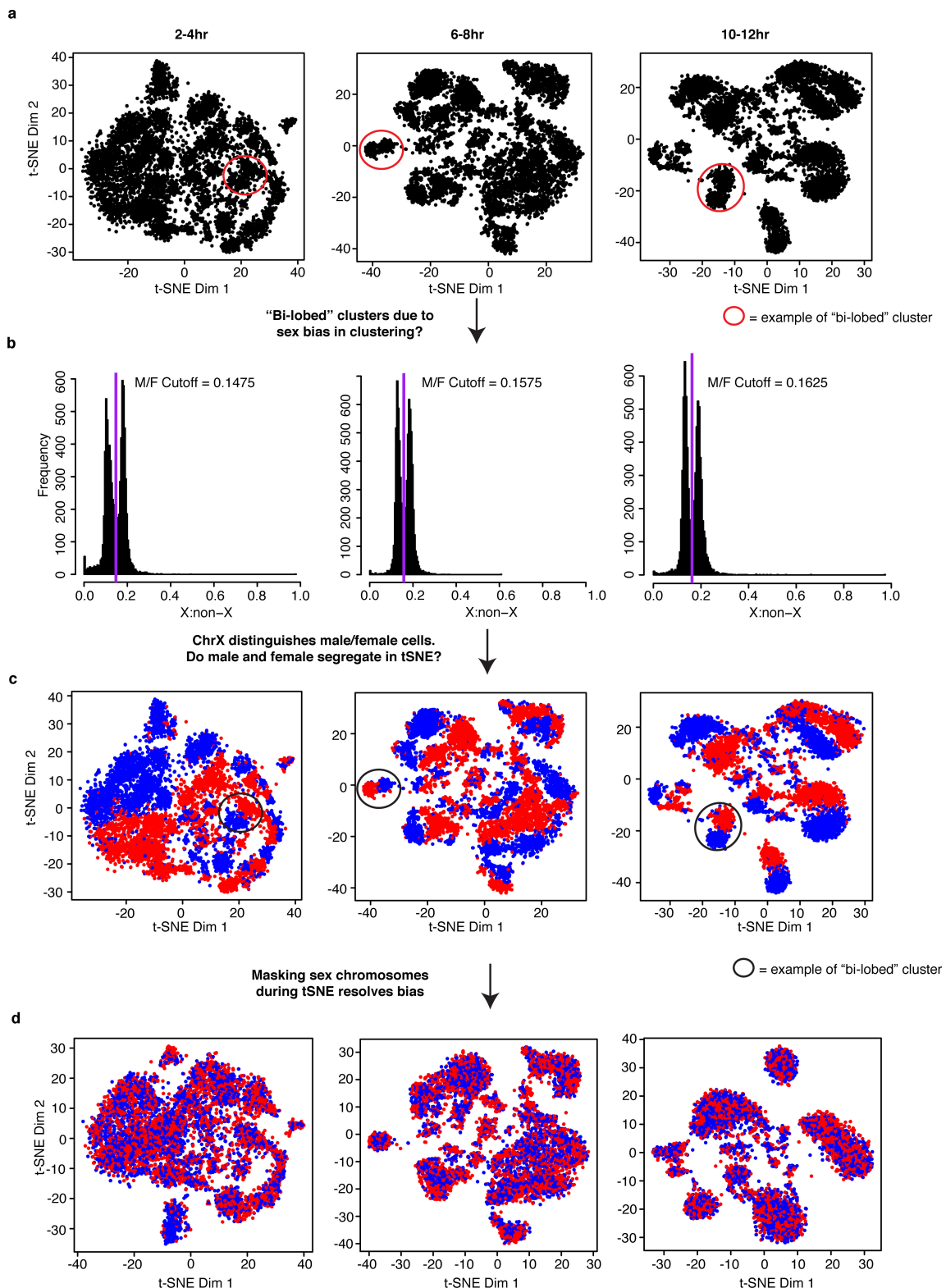
Extended Data Figure 3 | Relationship between transcription-factor binding motifs and occupancy, and LSI clade-specific accessibility. **a–c**, SeqGL was run on LSI clade-specific distal peaks at each time point to identify enriched sequence motifs. The top five most-enriched unique motifs for each clade are displayed. Coloured circles indicate the clade represented by each line. For the later time points (6–8 h and 10–12 h), blue is neurogenic ectoderm, yellow is non-neurogenic ectoderm, red is myogenic mesoderm and green is mesendoderm. The results show an enrichment of motifs for factors associated with early development at 2–4 h with more tissue-specific factor motifs (for example, mesodermal factor Mef2 or neural regulator Tramtrack) within germ-layer annotated

clades at later stages of development. **d–l**, Using ChIP occupancy data (peaks) and transcription factor binding motifs compiled previously¹⁶, we scanned for all transcription factor motif instances under ChIP peaks from datasets spanning 6–8 h of development using FIMO. Aggregate read counts in 4-kb windows centred on each identified motif instance are shown for each of the four LSI clades at 6–8 h. Green, endoderm; red, myogenic mesoderm; yellow, non-neurogenic ectoderm; blue, neurogenic ectoderm. Light shading in the same colours indicates 95% confidence intervals. **d–g**, Aggregate plots for four ubiquitous transcription factors (BEAF32, CTCF, Pho, and Trl) at 6–8 h. **h–l**, Aggregate plots for mesodermal transcription factors (Bap, Lmd, Mef2, Tin, Twi) at 6–8 h.



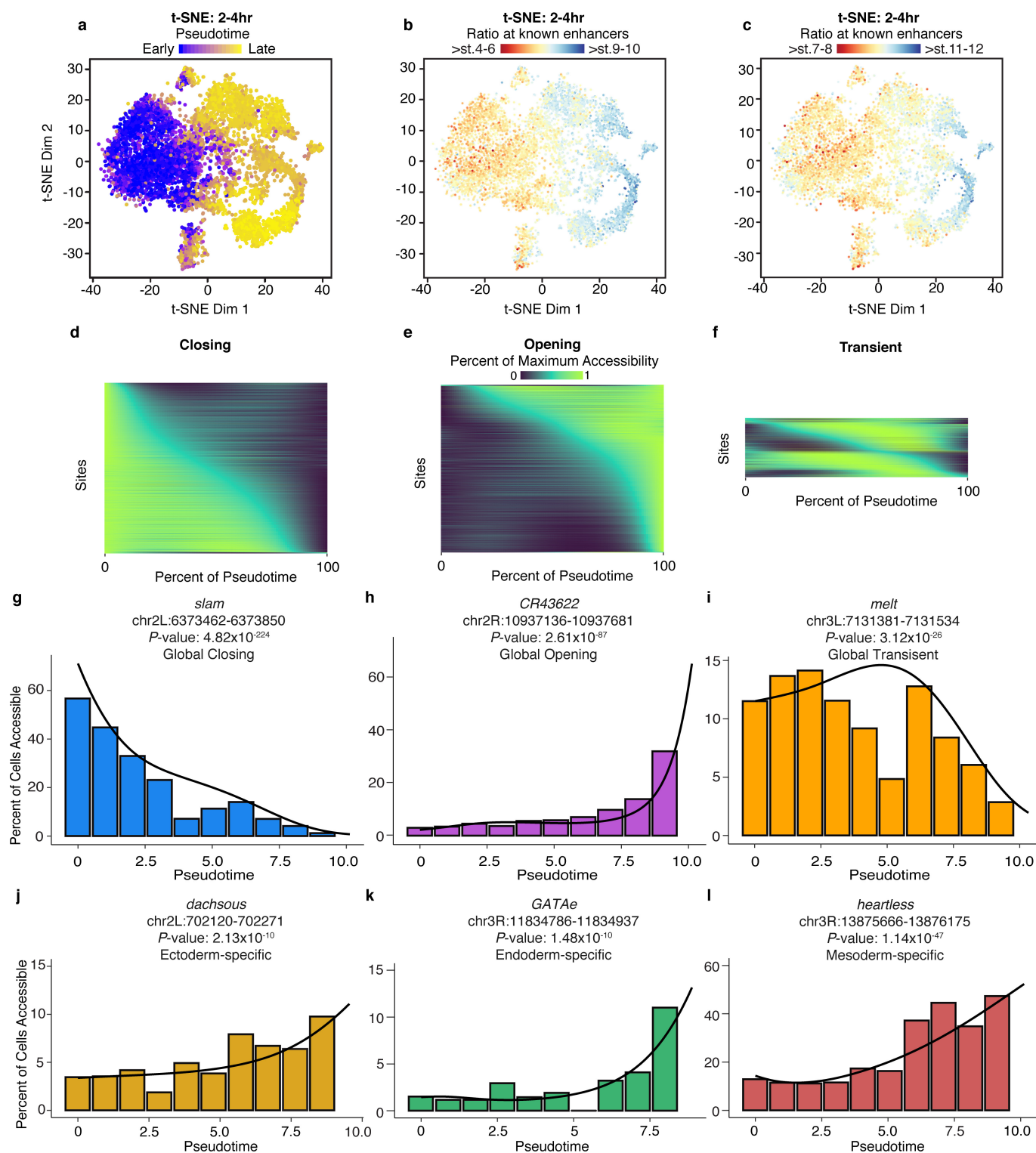
Extended Data Figure 4 | Similarities and differences in accessibility across all three time points. In addition to processing data from each time point independently, data from all cells can be analysed together (with the caveat that time point and batch are confounded). Here, we show binarized, LSI-transformed and clustered count data for 2-kb windows

across the genome for cells from all three time points (blue, 2–4 h; red, 6–8 h; orange, 10–12 h) processed together. The predominant pattern is one in which 2–4-h cells cluster separately from 6–8-h and 10–12-h cells. Cells from 6–8 h and 10–12 h are intermingled, clustering first (roughly) by germ layer of origin.



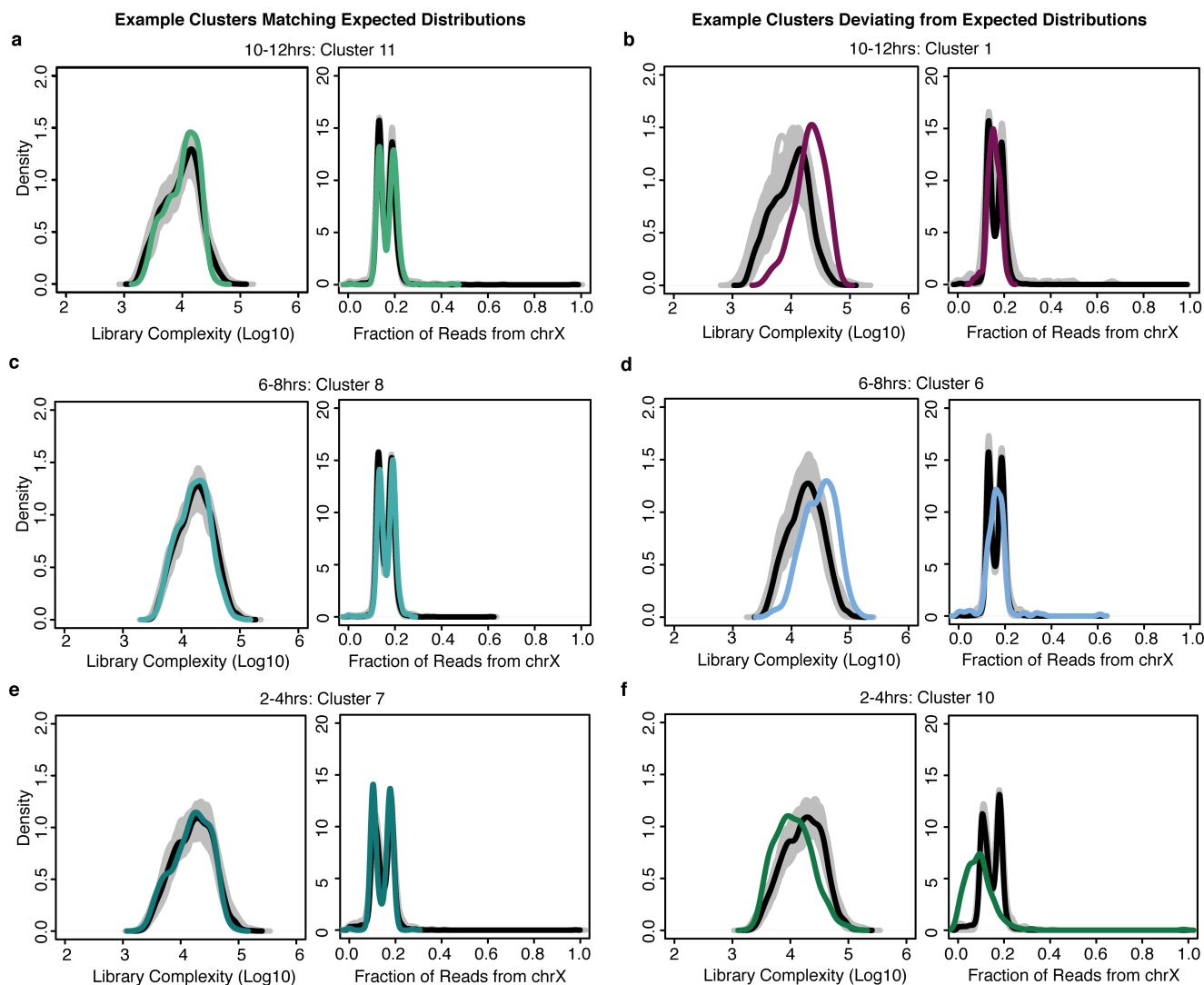
Extended Data Figure 5 | Sex of individual cells identified by ratio of X-chromosome to autosomal reads. Embryos at all stages consist of a mixture of male and female embryos (males, XY; females, XX). **a**, *t*-SNE plots of three time points from analysis in which sex chromosome sites were not excluded. Many clusters exhibited a bi-lobed structure, where each individual cluster was made up of two mirrored lobes (red circles identify one example of a bi-lobed cluster from each time point). This was most apparent at the 10–12-h time point. **b**, Histogram of the ratio of X-chromosome to autosomal reads in individual cells. To explore whether this bi-lobed structure was a function of sex biases in clustering, we attempted to sex individual cells. The ratio of X-chromosome to

autosomal reads shows a bimodal distribution, as expected in a system with heterogametic (XY) males and no evidence of imprinting. The purple line marks the local minimum between the two peaks of the histograms. **c**, Initial *t*-SNE clusters coloured according to sex assignment. Red indicates female cells and blue indicates male cells. Colouring individual cells by their sex reveals that the bi-lobed architecture is largely driven by sex biases in clustering. **d**, After removing X-chromosomal reads, data were re-clustered and individual cells were recoloured according to the ratio of X-chromosome to autosomal reads (red, female; blue, male). The resulting clusters showed an approximately equal number of male and female cells except for clusters 1 and 10 at the 2–4-h time point.



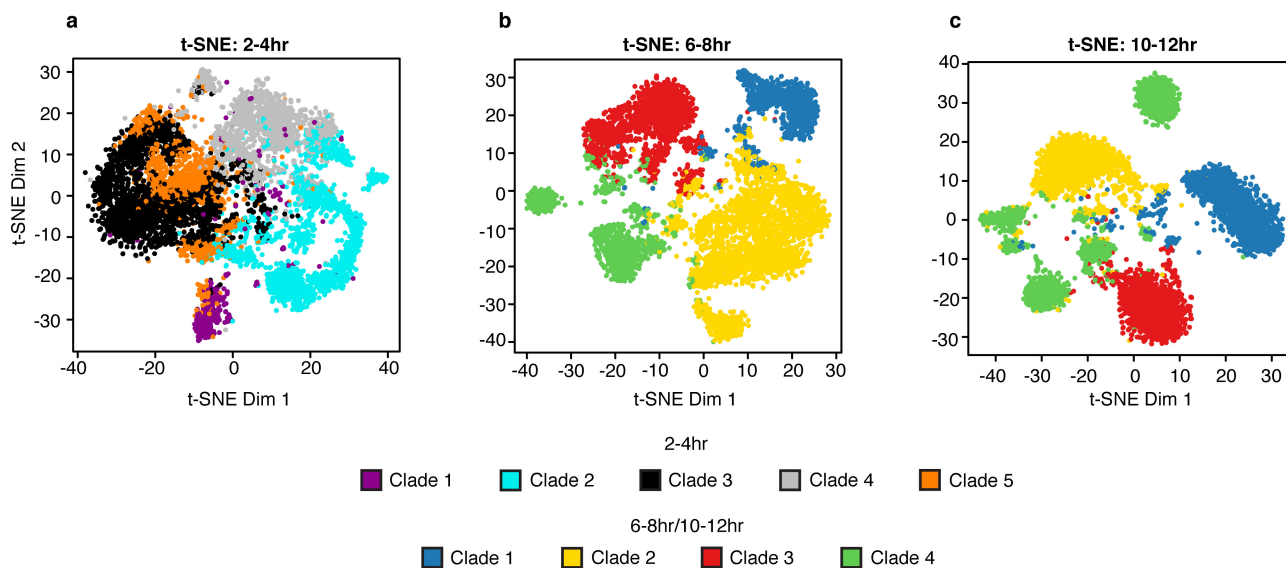
Extended Data Figure 6 | Temporal ordering of cells at 2–4 h using Monocle. **a–c**, t-SNE maps of cells at 2–4 h with colour representing either the Monocle-inferred pseudotime of each cell (**a**) or the ratio of reads per cell at enhancers active at different stages of development (**b**, **c**). Read counts within temporally characterized enhancers provide insight into the specific stage of development from which a cell is derived. Plotted here are ratios of counts in earlier versus later active enhancers showing a rough temporal progression from left to right that is also inferred by Monocle. **d–f**, Heat maps of sites that are significantly associated with pseudotime (based on a likelihood ratio test). For each site, a spline was fitted to the data across pseudotime. Sites (rows) were ordered for the heat maps based on the pseudotime at which they first reached half the maximum predicted accessibility from the fit curve. The colours indicate the spline-predicted

accessibility across pseudotime, scaled as the fraction of the maximum accessibility for that site. **g–i**, Single-locus plots of the most significant closing, opening and transient sites. Histogram of percentage of cells in which the specified site is accessible in 10 bins across pseudotime, within the 2–4 h time point. The curve is from spline fit for accessibility in cells through pseudotime. **j–l**, As in **g–i**, examples of sites with lineage-specific association with pseudotime. One example of a branch-specific opening site for each germ layer: ectoderm (**j**), endoderm (**k**) and mesoderm (**l**). In **g–i**, P values were calculated for likelihood ratio tests evaluating the effect of progress through pseudotime on accessibility ($n = 100$ bins of cells; see Methods for details). Note that the branch point in pseudotime occurs at approximately 5.6 on the x axis.



Extended Data Figure 7 | Library complexity and fraction of X chromosome reads highlights clusters of collisions between cells from different tissues. Density plots of the estimated library complexity (using the same equation implemented in Picard; left) and the representation of X-chromosome reads (right) in individual clusters. While most of the clusters defined by *t*-SNE are readily biologically interpretable, a small number of clusters (containing relatively few cells) were not easily characterized and are marked by an increase in both estimated library complexity and an unusual distribution of X chromosome to autosomal reads. These clusters are likely to be clusters of collisions; that is, cases in which two or more distinct cells share the same barcode as a consequence of the combinatorial indexing protocol. The black line is the global distribution for all cells in that time point. The grey lines show the

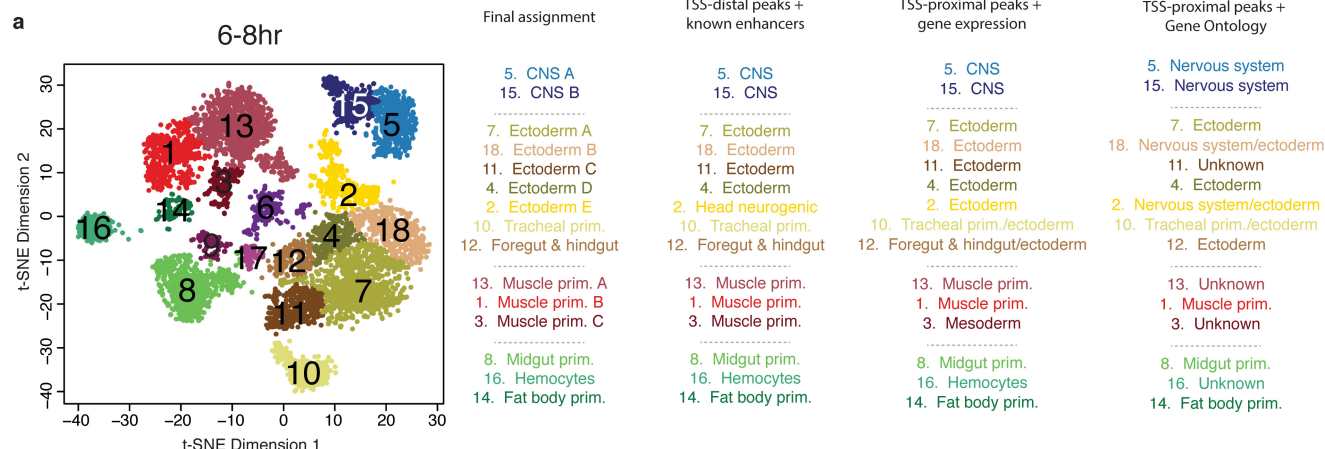
results of randomly sampling an equal number of cells to the cluster in question. The coloured line marks the distribution for the cluster being interrogated. **a, c, e**, Most clusters show relatively similar distributions of library complexity (left) and a characteristic, bimodal distribution among cells in the ratio of X-chromosome to autosomal reads (reflecting our use of a pool of male (XY) and female (XX) embryos, right). **b, d**, Putative collision clusters show a clear increase in the average library complexity (left) and a unimodal rather than bimodal distribution of X-chromosome to autosomal reads (right). **f**, These features are not universally diagnostic (for example, cluster 10 at 2–4 h seems to show a strong, bona fide sex bias), but the combination of features is strongly predictive of clusters containing few cells and conflicting biological annotations based on gene or enhancer overlaps.



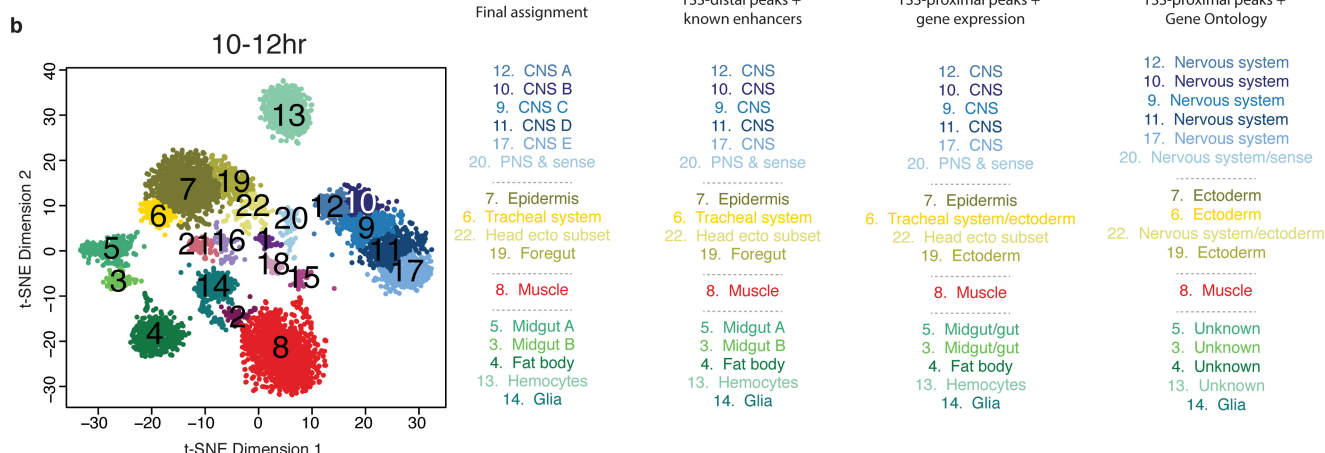
Extended Data Figure 8 | LSI-defined clades and *t*-SNE clusters show strong correspondence. *t*-SNE maps of cells from each of the three time points coloured according to the LSI clade to which they were previously assigned (Fig. 1a–c). For the post-gastrulation time points, green is endoderm, red is myogenic mesoderm, yellow is non-

neurogenic ectoderm and blue is neurogenic ectoderm. There is strong correspondence between the germ-layer-level clade annotations from the LSI analysis and tissue-specific *t*-SNE clusters, particularly at the post-gastrulation time points (6–8 h and 10–12 h).

Cluster assignment



Cluster assignment



Extended Data Figure 9 | Cell cluster assignment is similar using either enhancer or gene tissue activity. For each time point, cell clusters were annotated by first dividing peaks into TSS-distal (putative enhancers) and TSS-proximal (gene promoters) peaks. Each cell cluster was then annotated separately by overlaps between cluster-enriched peaks and: (1) enhancers, comparing the TSS-distal elements to the tissue or cell-type

activity of characterized enhancers; (2) genes, comparing TSS-proximal elements to the tissue expression of genes; and (3) Gene Ontology (GO) information (see Methods). Shown are the cluster assignments based on enhancer, gene expression, or Gene Ontology annotation alone. The final assignment, used in the main figures, combines all enrichment information to produce more robust assignments.

CORRIGENDUM

doi:10.1038/nature25991

Corrigendum: Cloche is a bHLH-PAS transcription factor that drives haemato-vascular specification

Sven Reischauer, Oliver A. Stone, Alethia Villasenor, Neil Chi, Suk-Won Jin, Marcel Martin, Miler T. Lee, Nana Fukuda, Michele Marass, Alec Witty, Ian Fiddes, Taiyi Kuo, Won-Suk Chung, Sherveen Salek, Robert Lerrigo, Jessica Alsiö, Shujun Luo, Dominika Tworus, Sruthy M. Augustine, Sophie Mucenieks, Björn Nystedt, Antonio J. Giraldez, Gary P. Schroth, Olov Andersson & Didier Y. R. Stainier

Nature **535**, 294–298 (2016); doi:10.1038/nature18614

In the Author Information section of this Letter, LT571435 from the SRA was provided as the accession number for the PacBio whole-genome sequencing data. Immediately after publication, we noted that LT571435 is instead the GenBank entry for the assembled genomic sequence of the *cloche* locus and this information is now included in the legend of Extended Data Fig. 2. The correct accession number for the PacBio whole-genome sequencing data, available from the European Nucleotide Archive, is PRJEB13442. The original Letter has been corrected online.

CORRECTIONS & AMENDMENTS

ERRATUM

doi:10.1038/nature25994

Erratum: Skin microbiota–host interactions

Y. Erin Chen, Michael A. Fischbach & Yasmine Belkaid

Nature **553**, 427–436 (2018); doi:10.1038/nature25177

Owing to a technical error, authors Y. Erin Chen and Michael A. Fischbach were both listed as corresponding authors instead of authors Michael A. Fischbach and Yasmine Belkaid in the HTML version of this Review (the PDF version was correct). This has been corrected online.

ERRATUM

doi:10.1038/nature26164

Erratum: The Beaker phenomenon and the genomic transformation of northwest Europe

Iñigo Olalde, Selina Brace, Morten E. Allentoft, Ian Armit, Kristian Kristiansen, Thomas Booth, Nadin Rohland, Swapan Mallick, Anna Szécsényi-Nagy, Alissa Mittnik, Eveline Altena, Mark Lipson, Iosif Lazaridis, Thomas K. Harper, Nick Patterson, Nasreen Broomandkhoshbacht, Yoan Diekmann, Zuzana Faltyskova, Daniel Fernandes, Matthew Ferry, Eadaoin Harney, Peter de Knijff, Megan Michel, Jonas Oppenheimer, Kristin Stewardson, Alistair Barclay, Kurt Werner Alt, Corina Liesau, Patricia Ríos, Concepción Blasco, Jorge Vega Miguel, Roberto Menduñía García, Azucena Avilés Fernández, Eszter Bánffy, Maria Bernabò-Brea, David Billoin, Clive Bonsall, Laura Bonsall, Tim Allen, Lindsey Büster, Sophie Carver, Laura Castells Navarro, Oliver E. Craig, Gordon T. Cook, Barry Cunliffe, Anthony Denaire, Kirsten Egging Dinwiddy, Natasha Dodwell, Michal Ernée, Christopher Evans, Milan Kuchařík, Joan Francès Farré, Chris Fowler, Michiel Gazenbeek, Rafael Garrido Pena, María Haber-Urriarte, Elżbieta Haduch, Gill Hey, Nick Jowett, Timothy Knowles, Ken Massy, Saskia Pfrengle, Philippe Lefranc, Olivier Lemerrier, Arnaud Lefebvre, César Heras Martínez, Virginia Galera Olmo, Ana Bastida Ramírez, Joaquín Lomba Maurandi, Tona Majó, Jacqueline I. McKinley, Kathleen McSweeney, Balázs Gusztáv Mende, Alessandra Modi, Gabriella Kulcsár, Viktória Kiss, András Czene, Róbert Patay, Anna Endrődi, Kitti Köhler, Tamás Hajdu, Tamás Szeniczey, János Dani, Zsolt Bernert, Maya Hoole, Olivia Cheronet, Denise Keating, Petr Velemínský, Miroslav Dobeš, Francesca Candilio, Fraser Brown, Raúl Flores Fernández, Ana-Mercedes Herrero-Corral, Sebastiano Tusa, Emiliano Carnieri, Luigi Lentini, Antonella Valenti, Alessandro Zanini, Clive Waddington, Germán Delibes, Elisa Guerra-Doce, Benjamin Neil, Marcus Brittain, Mike Luke, Richard Mortimer, Jocelyne Desideri, Marie Besse, Günter Brücken, Mirosław Furmanek, Agata Haluszko, Maksym Mackiewicz, Artur Rapiński, Stephany Leach, Ignacio Soriano, Katina T. Lillios, João Luís Cardoso, Michael Parker Pearson, Piotr Włodarczyk, T. Douglas Price, Pilar Prieto, Pierre-Jérôme Rey, Roberto Risch, Manuel A. Rojo Guerra, Aurore Schmitt, Joël Serrallongue, Ana Maria Silva, Václav Smrčka, Luc Vergnaud, João Zilhão, David Caramelli, Thomas Higham, Mark G. Thomas, Douglas J. Kennett, Harry Fokkens, Volker Heyd, Alison Sheridan, Karl-Göran Sjögren, Philipp W. Stockhammer, Johannes Krause, Ron Pinhasi, Wolfgang Haak, Ian Barnes, Carles Lalueza-Fox & David Reich

Nature **555**, 190–196 (2018); doi:10.1038/nature25738

In this Article, the surname of author Alessandra Modi was incorrectly listed as ‘Mod’. This has been corrected online.

Kit^{cre} knock-in mice fail to fate-map cardiac stem cells

ARISING FROM J. H. van Berlo *et al.* *Nature* **509**, 337–341 (2014); doi:10.1038/nature13309

In a cell fate-mapping study¹ using a *cre*-knock-in (KI) into the *Kit* locus, and in two other studies that used a similar genetic approach to track the fate of cardiac stem/progenitor cells (CSCs)^{2,3}, the authors concluded that c-Kit^{pos} (also known as Kit^{pos}) cells only negligibly contributed to the generation of cardiomyocytes. These studies questioned our findings⁴ that tissue-specific c-Kit^{pos} CSCs are endogenous regenerative agents that are necessary and sufficient for cardiomyocyte regeneration/replenishment after injury. There is a Reply to this Comment by van Berlo, J. H. *et al.* *Nature* **555**, <http://doi.org/10.1038/nature25772> (2018).

For these differences to be resolved, it is necessary to confirm that the *Kit*^{cre}-KI approach¹ correctly identifies and fate-maps c-Kit^{pos} CSCs and/or investigate whether the insertion of *Kit*^{cre}-KI affects CSC biology and cardiomyogenic potential.

We used tamoxifen-inducible *Kit*^{CreER(T2)/+ (5–7)}, *Kit*^{MerCreMer/+} (hereafter *Kit*^{MCM/+})^{1,2} and constitutive *Kit*^{CreGFPnls/+} (ref. 1) mouse lines, which are phenotypically similar to *Kit*^{W/+} mice⁵. These mice have white spotting in their fur coat, a 50% decrease in c-Kit expression (Extended Data Fig. 1a) and a testis growth deficit (with low fertility) when heterozygous, this genotype shows fetal/postnatal lethality in homozygotes^{1–3,6–8}.

The efficiency of recombination by the Cre-loxP system is proportional to the level of Cre expression and the duration of Cre expression in each cell^{9–11}, which in these *Kit*^{cre}-KI lines depends on the endogenous *Kit* promoter. In mice, different c-Kit^{pos} cell types express different levels of c-Kit, with mast cells being the highest c-Kit-expressing cells^{5–7}. In the adult mouse heart, the majority of c-Kit^{pos} cells (≥90%) are committed to the blood cell lineage, expressing markers such as CD45 and CD31 (and are lineage positive (Lin^{pos})) (Fig. 1a). Only a minority (<10%) of c-Kit^{pos} cardiac cells are CD45^{neg}CD31^{neg} (Lin^{neg}) (Fig. 1a). These Lin^{neg}c-Kit^{pos} cardiac cells are enriched for and include all of the adult CSCs, which comprise only approximately 10% of these cells^{12,13}. This subset expresses low but clearly detectable levels of *Kit* mRNA and c-Kit protein (Extended Data Fig. 1b), which are significantly lower than the levels found in Lin^{pos}c-Kit^{pos} cardiac cells (Extended Data Fig. 1b). Single wild-type Lin^{neg}c-Kit^{pos} clonogenic and multipotent CSCs show significantly lower *Kit* expression, at the mRNA and protein levels, than embryonic stem cells, haematopoietic stem cells (HSCs) and bone marrow mast cells (Extended Data Fig. 1c, d). Accordingly, Cre^{ER(T2)} protein and mRNA expression in bone marrow mast cells is robust, whereas it is only faintly detectable in freshly isolated Lin^{neg} CSCs (Extended Data Fig. 1e, f). This heterogeneity in cell types that have different levels of c-Kit expression, the low abundance of CSCs among the c-Kit^{pos} cardiac cells and the very low levels of c-Kit expression in CSCs highlight the difficulties of using *Kit*^{cre}-KI to track the fate of the CSCs.

Kit^{CreER(T2)/+} mice^{6–8} were crossed with homozygous global double-fluorescent Cre-reporter *Rosa26*^{mT/mG} mice¹⁴ (Extended Data Fig. 2a), expressing a membrane-targeted tandem Tomato dimer (mT) that switches to a membrane-targeted GFP (mG) after Cre-dependent recombination¹⁴.

When double-mutant *Kit*^{CreER(T2)/+;Rosa26}^{mT/mG/+} mice were given a standard tamoxifen diet for 14 days, 80 ± 8% of the c-Kit^{pos} bone marrow mast cells showed Cre-dependent recombination and

expressed GFP, whereas less than 5% of HSCs showed recombination and GFP expression (Extended Data Fig. 1g–j). At the same time, ≤20% of all c-Kit^{pos} cardiac cells were recombined (Extended Data Fig. 1k), but all of these were CD45^{pos}, CD31^{pos} or both, representing cardiac mast cells or endothelial progenitor cells (Extended Data Fig. 1l). By contrast, the CSC-enriched Lin^{neg}c-Kit^{low} cells showed minimal recombination (≤1%) and expression of GFP (Extended Data Fig. 1l). Using the *Kit*^{MCM/+;Rosa26}^{mT/mG/+} mice from van Berlo *et al.*¹ yielded almost identical results (Extended Data Fig. 3a–c).

To increase the probability of recombination by extending the duration of Cre expression, 8-week-old double-mutant *Kit*^{CreER(T2)/+;Rosa26}^{mT/mG/+} mice received a tamoxifen diet for four months¹ (Fig. 1b–e). This regime showed efficient recombination in several tissue-specific c-Kit-expressing cells (Extended Data Fig. 2b). In the bone marrow, approximately 80% of the total number of cells and up to 60% of c-Kit^{pos} cells became GFP^{pos} (Extended Data Fig. 2c). Analysis of bone marrow cell sub-populations showed that approximately 100% of c-Kit^{pos} mast cells were GFP^{pos} (Extended Data Fig. 2d), whereas only ≤35% of HSCs expressed GFP (Extended Data Fig. 2e). As expected, recombination events were increased further by constitutive *Kit*^{cre} expression in *Kit*^{CreGFPnls/+;Rosa26}^{mT/mG/+} double-heterozygous mice (Extended Data Fig. 2g–j).

Fluorescence-activated cell sorting (FACS) of dissociated total cardiac cells showed that around 15% of cells showed Cre-mediated recombination and GFP expression (Fig. 1b). Among the total c-Kit antibody-labelled cells, 65 ± 5% were GFP^{pos} (Fig. 1c). Immunohistochemistry of frozen cardiac sections confirmed that approximately 60% of c-Kit antibody-labelled cells had turned GFP^{pos} (Extended Data Fig. 2f). The majority of these GFP^{pos} cells were still mT^{pos} (Fig. 1b, c), highlighting its slow decay in myocardial tissue¹⁴. Nevertheless, 80% of the endothelial/mast cell lineage-committed Lin^{pos}c-Kit^{pos} cells were GFP^{pos}, whereas only ≤10% of the CSC-enriched Lin^{neg}c-Kit^{low} cardiac cells showed recombination (Fig. 1d). Experiments using *Kit*^{MCM/+} mice¹ yielded similar findings (data not shown).

After mice were fed the tamoxifen diet for four months, only 0.04 ± 0.01% cardiomyocytes were GFP^{pos}, and most of those cells were also mT^{pos} (Fig. 1e). van Berlo *et al.*¹ interpreted the presence of double-labelled mT^{pos} and GFP^{pos} cardiomyocytes as the evidence of fusion between a pre-existing *Rosa26*^{mT/mG/+} cardiomyocyte and a *Kit*^{cre/+} cell. However, because myocardial decay of mT is very slow¹⁴, mT/mG double-positive cardiomyocytes are most likely the consequence of the very slow decay of mT in randomly and progressively recombined progenitors that differentiate into cardiomyocytes.

Similar results were obtained when using 8–12-week-old *Kit*^{CreGFPnls/+;Rosa26}^{mT/mG/+} double-heterozygous mice (Fig. 1f–i) and *Kit*^{CreGFPnls/+} mice crossed with *Rosa26*^{floxed-stop-tdTomato} mice (Extended Data Fig. 3d–f).

We sorted and cultured Lin^{neg}c-Kit^{pos} CSCs from *Kit*^{CreER(T2)/+;Rosa26}^{mT/mG/+} double-heterozygous mice in the presence or absence of tamoxifen *in vitro*. Tamoxifen induced recombination in ≤1% of the Lin^{neg}c-Kit^{pos} cells and expression of GFP (Fig. 1j), whereas >95% of bone-marrow-derived mast cells showed recombination within seven days (Fig. 1k). None of the single-cell-derived CSC clones that were generated in medium with tamoxifen at passage (P)4

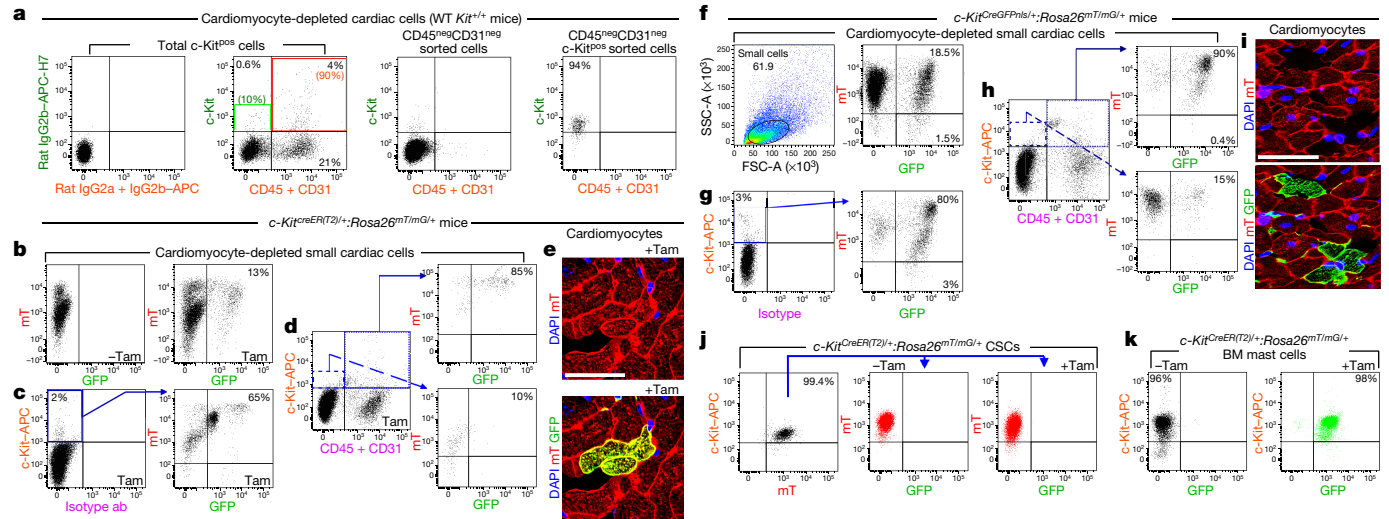


Figure 1 | *Kit^{cre}* genetic lineage cell tracking tags only a minimal fraction of CSCs *in vivo* and *in vitro*. **a**, Left, flow cytometry dot plots show that the majority of c-Kit^{pos} cardiac cells are also CD45^{pos} or CD31^{pos} (red box), whereas only a minority are CD45^{neg} and CD31^{neg} (green box). Right, flow cytometry plots show that CD45^{pos}CD31^{pos}(Lin^{pos}) cells are efficiently removed from cardiac cells by negative sorting and CD45^{neg}CD31^{neg}c-Kit^{pos}-sorted cells uniformly express low levels of c-Kit. The panels show an experiment with sequential analysis of unsorted cells (left) and re-analysis of sorted cells (right) from cardiomyocyte-depleted cells obtained from five enzyme-digested and pooled hearts. Representative of *n* = 3 experiments. **b**, Representative FACS plots showing that the tamoxifen (Tam) diet for four months induces GFP expression in 15% of all cardiac cells from *Kit^{CreER(T2)/+;Rosa26^{mT/mG}/+}* mice. No GFP signal is detected if the mice were not fed the tamoxifen diet. **c**, Approximately 70% of c-Kit-labelled cardiac cells were GFP^{pos} after the mice were fed the tamoxifen diet. Almost all of these GFP^{pos} cells were still mT^{pos}. **d**, The majority of c-Kit^{pos} cardiac cells are CD45^{pos} (labelled with CD45-PE-Vio) or CD31^{pos} (labelled with CD31-PE-Vio) and around 80% of them are recombined to express GFP, whereas $\leq 10\%$ of the CSC-enriched CD45^{neg}CD31^{neg}c-Kit^{pos} cardiac cells became GFP^{pos}. Almost

were GFP^{pos}. Additionally, when Lin^{neg}c-Kit^{pos} cardiac cells from *Kit^{CreGFPnl/+;Rosa26^{mT/mG}/+}* double-heterozygous mice were cloned, only 1 out of 19 CSC clones (from 1,056 deposited single cells, 116 of which were GFP^{pos}) was GFP^{pos}. When 10 randomly selected GFP^{neg} clones were expanded for up to 10 passages, only $8 \pm 3\%$ cells within each clone became GFP^{pos}.

Together, these data show that *Kit^{cre}* alleles, inducible and constitutive, in the cardiac Lin^{neg}c-Kit^{low} cell pool, which includes multipotent CSCs, recombine very inefficiently, stochastically and do not reliably track the fate of Lin^{neg}c-Kit^{low} CSCs.

Kit^{cre} KI null alleles are phenotypically equivalent to severe W mutations that affect the molecular and phenotypic properties of c-Kit-expressing cells^{5,15}. We therefore evaluated the growth, self-renewal and myogenicity of freshly isolated CSC-enriched Lin^{neg}c-Kit^{pos} cardiac cells and of single-cell-derived c-Kit^{pos} CSC clones obtained from *Kit^{CreER(T2)/+}* mice (hereafter, W^{Cre} CSCs) compared to freshly isolated and cloned *Kit^{+/+}* CSCs from wild-type littermates. Freshly isolated W^{Cre} CSCs at P1 showed a significant growth delay and lower BrdU incorporation in response to serum, and formed fewer single-cell-derived clones and cardiospheres compared to wild-type CSCs (Fig. 2a–c). Similar data were obtained with CSC-enriched Lin^{neg}c-Kit^{pos} cardiac cells from *Kit^{CreGFPnl/+}* mice (Extended Data Fig. 4a). When tested for cardiomyogenic differentiation^{12,13}, wild-type CSC-derived cardiospheres homogeneously differentiated into cardiac troponin I

all of these GFP^{pos} cells were still mT^{pos}. **b–d**, Representative of *n* = 5 hearts. **e**, Confocal image showing recombined GFP^{pos} cardiomyocytes after four months of tamoxifen diet in *Kit^{CreER(T2)/+;Rosa26^{mT/mG}/+}* mice. These cardiomyocytes still express mT. Representative of *n* = 5 hearts. Scale bar, 20 μ m. **f**, Approximately 20% of all cardiac cells of 8–12-week-old *Kit^{CreGFPnl/+;Rosa26^{mT/mG}/+}* mice are GFP^{pos}. **g**, Approximately 80% of c-Kit^{pos} cardiac cells are GFP^{pos} from same mice. **h**, Around 90% of endothelial/mast cell lineage-committed, CD45^{pos}CD31^{pos}c-Kit^{pos} cardiac cells are recombined to express GFP, whereas fewer than 20% of the CSC-enriched CD45^{neg}CD31^{neg}c-Kit^{pos} were recombined to become GFP^{pos}. **f–h**, Representative of *n* = 5 hearts. Note that GFP-expressing cells in **f–h** are mostly still mT^{pos}. **i**, Confocal image showing recombined GFP^{pos} cardiomyocytes (around 0.25% of all cardiomyocytes) in 8–12-week-old *Kit^{CreGFPnl/+;Rosa26^{mT/mG}/+}* mice. Representative of *n* = 5 hearts. Scale bar, 20 μ m. Most of these cardiomyocytes do not express mT. **j**, Representative flow cytometry plots showing that tamoxifen *in vitro* is unable to recombine c-Kit^{pos} CSCs isolated from *Kit^{CreER(T2)/+;Rosa26^{mT/mG}/+}* mice. *n* = 5 biological replicates. **k**, Tamoxifen addition *in vitro* induced recombination in all bone-marrow (BM)-derived mast cells. *n* = 5 biological replicates.

(cTnI)-expressing cells (Fig. 2d), exhibited vigorous spontaneous rhythmic contractions (Supplementary Video 1), and showed upregulation of the main cardiomyocyte-specific genes (Fig. 2e). By contrast, cardiospheres from W^{Cre} CSCs rarely differentiated into cTnI-positive cells and sparingly showed rare/weak beating (Fig. 2d). Accordingly, cardiac-specific gene transcripts were minimally upregulated (Fig. 2e). RNA-sequencing-based analysis of whole-cell transcriptomes of cloned W^{Cre} CSCs and wild-type CSCs confirmed that c-Kit heterozygosity affected the expression of genes associated with RNA transcription, cardiac development, and cell growth, survival and death (Extended Data Fig. 4b, c).

To test whether restoring normal levels of c-Kit expression would correct defects in the stemness potential of W^{Cre} CSCs *in vitro* and *in vivo*, we transfected a bacterial artificial chromosome (BAC) construct spanning the entire *Kit* gene locus into cloned mT-W^{Cre} CSCs (from *Kit^{CreER(T2)/+;Rosa26^{mT/mG}/+}* mice). Using qPCR, we selected one clone with a single BAC^{Kit} copy (hereafter mT-BAC^{Kit}W^{Cre} CSCs) that expressed *Kit* mRNA and c-Kit protein at levels similar to those of mT wild-type (mT-WT) CSCs (from *Rosa26^{mT/mG}/+* littermates) (Fig. 2f and Extended Data Fig. 4d, e). The defects in cell growth, clonogenesis, cardiosphere formation and myogenic differentiation of mT-W^{Cre} CSCs were completely rescued by BAC^{Kit} transfection. The mT-BAC^{Kit}W^{Cre} CSCs phenotype is indistinguishable from that of mT-WT CSCs (Fig. 2g, h and Extended Data Fig. 4f–h).

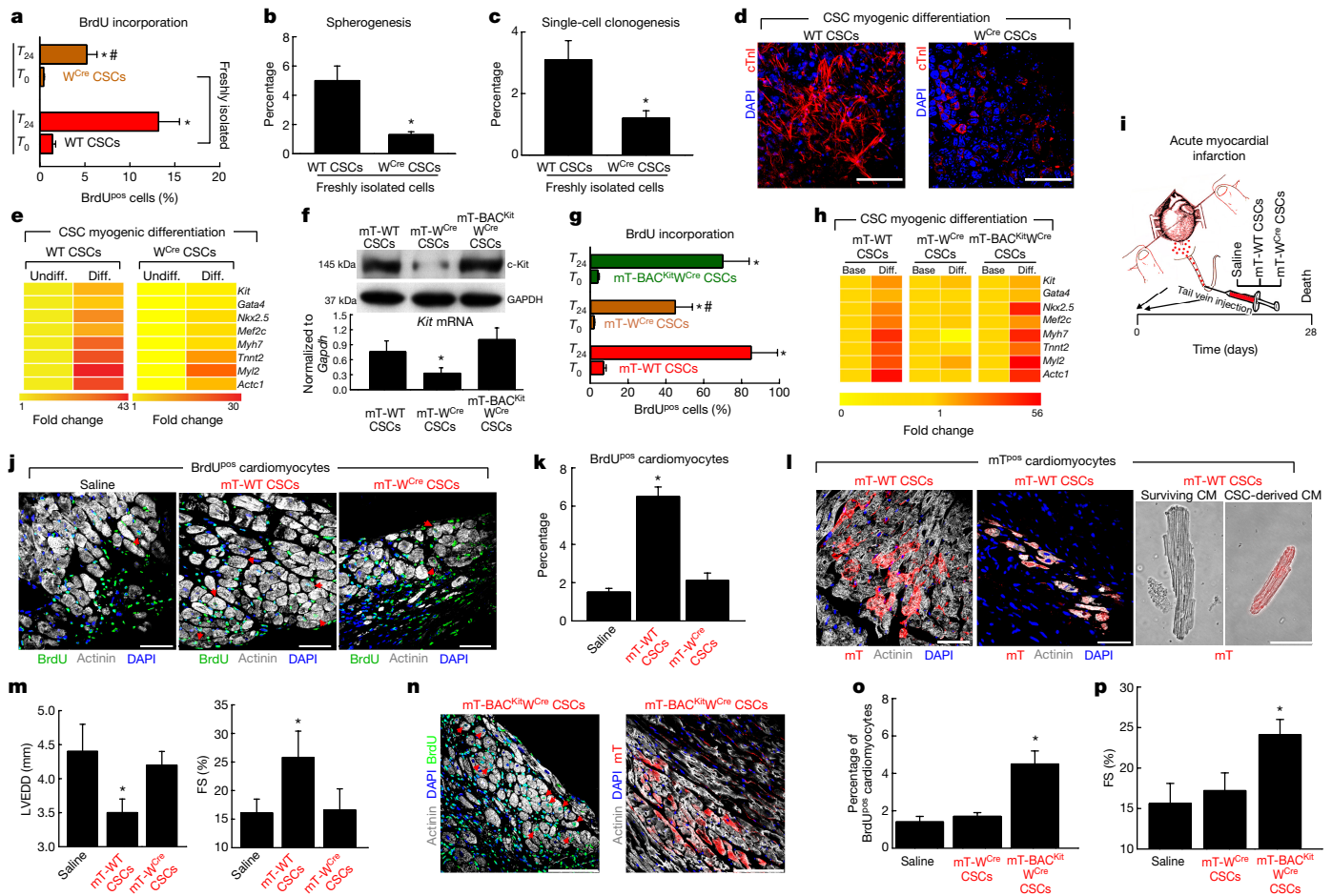


Figure 2 | The W^{Cre} allele impairs c-Kit^{pos} adult CSC myogenic and regenerative properties. **a**, Serum-induced 24 h BrdU incorporation in freshly isolated W^{Cre} CSCs compared to wild-type (WT) CSCs. $*P < 0.05$ vs. T_0 (before addition); $\#P < 0.05$ compared to wild-type CSCs. **b**, W^{Cre} CSCs formed fewer cardiospheres than wild-type CSCs. 1×10^5 plated cells. **c**, Clonal efficiency of freshly isolated W^{Cre} CSCs versus wild-type CSCs. **b**, **c**, $*P < 0.05$ versus wild-type CSCs. **d**, cTnI expression in cardiospheres from cloned wild-type CSCs and W^{Cre} CSCs after 14 days in myogenic medium. Scale bar, 50 μ m. **a–d**, Representative of $n = 5$ biological replicates. **e**, Heat map showing qPCR analysis of main cardiomyocyte genes in cardiospheres from cloned wild-type and W^{Cre} CSCs after 14 days in myogenic medium. $n = 3$ biological replicates. **f**, Normalization of *Kit* mRNA and protein expression in mT-BAC^{Kit} W^{Cre} CSCs compared to mT-WT CSCs and naive mT- W^{Cre} CSCs. $*P < 0.05$ compared to all other treatments. $n = 3$ biological replicates. **g**, BrdU incorporation in mT-BAC^{Kit} W^{Cre} CSCs compared to mT- W^{Cre} CSCs and mT-WT CSCs. $*P < 0.05$ versus T_0 ; $\#P < 0.05$ versus T_{24} . $n = 4$ biological replicates. **h**, Heat map of qPCR analysis of the modulation of main cardiomyocyte genes in cardiospheres from mT-BAC^{Kit} W^{Cre} CSCs compared to mT- W^{Cre} CSCs and mT-WT CSCs after 14 days in myogenic medium. Representative of $n = 3$ biological replicates. **i**, Myocardial infarction (MI) study design. **j**, BrdU incorporation (red arrows point at BrdU^{pos} cardiomyocytes) in the border zone of infarcted hearts from wild-type (WT) mice treated with either saline ($n = 6$) or with

the indicated CSC types. Scale bar, 50 μ m. **k**, Newly generated BrdU^{pos} cardiomyocytes 28 days after myocardial infarction in wild-type mice treated with either saline or with the indicated CSC types. $*P < 0.05$ versus all other treatments. **j–m**, $n = 6$ mice per group. **l**, Left, CSC-derived mT^{pos} cardiomyocytes in the border zone of infarcted hearts from mice treated with mT-WT CSCs. Note that, in the middle panel, mT^{pos} cardiomyocytes remuscularize proper infarct zone. Right, single cardiomyocytes dissociated from the hearts of additional three mice with a myocardial infarction. A living mT-fluorescing CSC-derived bona fide cardiomyocyte (CM) is shown opposite a host surviving cardiomyocyte. Scale bar, 50 μ m. **m**, Left ventricular function 28 days after myocardial infarction in wild-type mice treated with either saline or with the indicated CSCs types. $*P < 0.05$ versus saline and mT- W^{Cre} CSCs. LVEDD, left ventricle end diastolic diameter; FS, fractional shortening. **j**, **k**, **m**, $n = 6$ mice per group. **n**, BrdU^{pos} cardiomyocytes (red arrows) in the border zone of infarcted hearts from wild-type mice treated with mT-BAC^{Kit} W^{Cre} CSCs. Scale bar, 50 μ m. **o**, Percentage of newly generated BrdU^{pos} cardiomyocytes 28 days after myocardial infarction in wild-type mice treated with saline ($n = 6$), mT- W^{Cre} CSCs ($n = 6$) or mT-BAC^{Kit} W^{Cre} CSCs ($n = 6$). $*P < 0.05$ versus all other treatments. **p**, Echocardiography assessment of left ventricular function (fractional shortening) 28 days after myocardial infarction in wild-type mice treated with either saline or the indicated CSC types. $*P < 0.05$ versus all other treatments. n values as in **o**. All data are mean \pm s.d.

To test the effect of the *Kit*^{Cre}-KI mutations on cardiac repair/regeneration, *Kit*^{CreGFPnls/+};*Rosa26*^{mT/mG/+} double-heterozygous mice and *Rosa26*^{mT/mG/+} heterozygous controls were subjected to myocardial infarction followed by systemic BrdU administration. At 28 days after myocardial infarction, *Kit*^{CreGFPnls/+};*Rosa26*^{mT/mG/+}

mice had significantly larger infarcts and lower number of BrdU^{pos} cardiomyocytes in the infarct border zone than *Rosa26*^{mT/mG/+} heterozygous controls (0.4 ± 0.14 compared to $1.1 \pm 0.15\%$, respectively; $P < 0.05$) (Extended Data Fig. 5a–d). The number of c-Kit^{pos} cell-generated GFP^{pos} cardiomyocytes in *Kit*^{CreGFPnls/+};*Rosa26*^{mT/mG/+}

mice after myocardial infarction was only slightly higher than in the sham-operated controls ($0.28 \pm 0.04\%$ compared to $0.4 \pm 0.2\%$, respectively; $P > 0.05$, not significant) (Extended Data Fig. 5e–g). The lower number of newly generated GFP^{pos} or BrdU^{pos} cardiomyocytes in *Kit*^{CreGFPnls/+};*Rosa26*^{mT/mG/+} compared to *Rosa26*^{mT/mG/+} mice correlated inversely with increased hypertrophy of the residual cardiomyocytes (Extended Data Fig. 5h, i) and reduced arteriolar and capillary density (Extended Data Fig. 5j, k) in the infarct border zone. Concordantly, left ventricular dysfunction was more severe in *Kit*^{CreGFPnls/+};*Rosa26*^{mT/mG/+} than in *Rosa26*^{mT/mG/+} mice (Extended Data Fig. 5l, n). These data demonstrate that the *W*^{Cre} allele negatively affects the myogenicity of the CSCs, but cannot address whether the latter is the main mechanism that underlies the worsened cardiac remodelling after myocardial infarction in *Kit*^{Cre} mice. This is because c-Kit haploinsufficiency could also affect additional repair/regeneration processes other than cardiomyocyte replacement.

We next transplanted 4×10^5 cloned *W*^{Cre} CSCs obtained from *Kit*^{CreERT2/+};*Rosa26*^{mT/mG/+} mice (hereafter mT-*W*^{Cre} CSCs) or the same number of wild-type CSCs from *Rosa26*^{mT/mG/+} littermates (mT-WT CSCs) or saline into the systemic circulation 30 min after myocardial infarction in three groups of C57BL/6J mice (Fig. 2i). After 28 days, recipients of mT-WT CSCs had a significant higher number of BrdU^{pos} cardiomyocytes in the infarct border zone than the mT-*W*^{Cre} CSCs- and saline-injected mice ($6.5 \pm 0.5\%$ compared to $2.0 \pm 0.4\%$ and $1.5 \pm 0.2\%$, respectively; $P < 0.05$) (Fig. 2j, k). The mT-WT CSCs generated a significant number of mT^{pos} cardiomyocytes within the infarct border zone ($3.5 \pm 0.5\%$) with foci of re-muscularized tissue within the proper infarct zone (Fig. 2l). This result was confirmed by identifying abundant mature mT^{pos} cardiomyocytes in dissociated myocardial cells (Fig. 2l) together with a notable improvement in arteriole and capillary density in the mT-WT CSC-injected mice. By contrast, only very small numbers of mT^{pos} cardiomyocytes with scarce arterioles and capillaries were detected in the infarct border zone of the mT-*W*^{Cre} CSC-injected mice (Extended Data Fig. 6a–d). Not surprisingly, whereas mT-WT CSCs significantly improved left ventricular function over saline injection, mT-*W*^{Cre} CSCs did not (Fig. 2m).

In concordance with the results presented above, BAC^{Kit} transgenesis also completely rescued regenerative defects of the *W*^{Cre} CSCs (Fig. 2n–p and Extended Data Fig. 6e–g). The beneficial effects of mT-BAC^{Kit}*W*^{Cre} CSCs transplantation on myocardial remodelling and function after myocardial infarction (Fig. 2n–p and Extended Data Fig. 6e–g) were qualitatively and quantitatively similar to those produced by mT wild-type CSCs (see Fig. 2j–m and Extended Data Fig. 6a–d).

To directly test whether cell fusion, not CSC-derived neomyogenesis (see ref. 1), causes the appearance of double-labelled mT/mG cardiomyocytes after wild-type CSC transplantation, mT-WT CSCs from *Rosa26*^{mT/mG} mice (mT^{pos}) were injected into *Tg*^{Myh6-MCM} mice after myocardial infarction. Mice were then fed with the tamoxifen diet for 28 days. Invariably, all the mT^{pos} cardiomyocytes were only mT^{pos} and none expressed GFP (see Extended Data Fig. 6h–j).

These data show that harvested endogenous CSCs are robustly myogenic *in vitro* and *in vivo* and, while they do not further address whether and/or to what extent c-Kit^{pos} CSCs contribute to endogenous cardiac regeneration (but see ref. 4), they do reveal several limitations of the use of *Kit*^{Cre}-KI strategies for CSC identification and cell-fate mapping. The very low number of endogenous c-Kit^{pos} CSC-generated cardiomyocytes detected in the *Kit*^{Cre} mice simply reflects the failure to recombine the CSCs to track their progeny and the severe defect in CSC myogenesis produced by the *Kit*^{Cre} allele.

This work was supported by grants from CARE-MI FP7 (Health-F5-2010-242038) large-scale collaborative project, M.I.U.R. FIRB-Futuro-in-Ricerca (RBF1213KA), M.I.U.R. PRIN2015

(2015ZTT5KB_004), and the Italian Ministry of Health (GR-2010-2318945).

Data Availability RNA-sequencing data are available in GEO with the accession number GSE102002.

Carla Vicinanza^{1*}, Iolanda Aquila^{1*}, Eleonora Cianflone¹, Mariangela Scalise¹, Fabiola Dorino¹, Teresa Mancuso¹, Francesca Fumagalli², Emilia Dora Giovannone³, Francesca Cristiano⁴, Enrico Iaccino¹, Pina Marotta¹, Annalaura Torella⁵, Roberto Latini², Valter Agosti³, Pierangelo Veltri⁴, Konrad Urbanek⁶, Andrea M. Isidori⁷, Dieter Saur^{8,9}, Ciro Indolfi¹, Bernardo Nadal-Ginard^{1§} & Daniele Torella^{1§}

¹Molecular and Cellular Cardiology, Department of Medical and Surgical Sciences, Magna Graecia University, Catanzaro 88100, Italy.

emails: bernardo.nadalginard@gmail.com; dtorella@unicz.it

²IRCCS - Istituto di Ricerche Farmacologiche “Mario Negri”, Milan 20156, Italy.

³CIS for Genomics and Molecular Pathology, Magna Graecia University, Catanzaro 88100, Italy.

⁴Bioinformatics, Department of Medical and Surgical Sciences, Magna Graecia University, Catanzaro 88100, Italy.

⁵Department of Biochemistry, Biophysics and General Pathology, University of Campania “L. Vanvitelli”, Naples 80121, Italy.

⁶Department of Experimental Medicine, Section of Pharmacology, University of Campania “L. Vanvitelli”, Naples 80121, Italy.

⁷Department of Experimental Medicine, “La Sapienza” University, Rome 00161, Italy.

⁸Division of Translational Cancer Research, German Cancer Research Center (DKFZ) and German Cancer Consortium (DKTK), 69120 Heidelberg, Germany.

⁹Chair of Translational Cancer Research and Department of Medicine II, School of Medicine, Klinikum rechts der Isar, Technische Universität München, 81675 München, Germany.

*These authors contributed equally to this work.

§These authors jointly supervised this work.

Received 10 May; accepted 15 November 2017.

- van Berlo, J. H. *et al.* c-kit⁺ cells minimally contribute cardiomyocytes to the heart. *Nature* **509**, 337–341 (2014).
- Sultana, N. *et al.* Resident c-kit⁺ cells in the heart are not cardiac stem cells. *Nat. Commun.* **6**, 8701 (2015).
- Liu, Q. *et al.* Genetic lineage tracing identifies *in situ* Kit-expressing cardiomyocytes. *Cell Res.* **26**, 119–130 (2016).
- Ellison, G. M. *et al.* Adult c-kit^{pos} cardiac stem cells are necessary and sufficient for functional cardiac regeneration and repair. *Cell* **154**, 827–842 (2013).
- Lennartsson, J. & Rönstrand, L. Stem cell factor receptor/c-Kit: from basic science to clinical implications. *Physiol. Rev.* **92**, 1619–1649 (2012).
- Klein, S. *et al.* Interstitial cells of Cajal integrate excitatory and inhibitory neurotransmission with intestinal slow-wave activity. *Nat. Commun.* **4**, 1630 (2013).
- Heger, K. *et al.* CreERT2 expression from within the c-Kit gene locus allows efficient inducible gene targeting in and ablation of mast cells. *Eur. J. Immunol.* **44**, 296–306 (2014).
- Hatzistergos, K. E. *et al.* cKit⁺ cardiac progenitors of neural crest origin. *Proc. Natl Acad. Sci. USA* **112**, 13051–13056 (2015).
- Jensen, P. & Dymecki, S. M. Essentials of recombinase-based genetic fate mapping in mice. *Methods Mol. Biol.* **1092**, 437–454 (2014).
- Aquila, I. *et al.* The use and abuse of Cre/Lox recombination to identify adult cardiomyocyte renewal rate and origin. *Pharmacol. Res.* **127**, 116–128 (2018).
- Nadal-Ginard, B., Ellison, G. M. & Torella, D. Absence of evidence is not evidence of absence: pitfalls of cre knock-ins in the c-Kit locus. *Circ. Res.* **115**, 415–418 (2014).
- Smith, A. J. *et al.* Isolation and characterization of resident endogenous c-Kit⁺ cardiac stem cells from the adult mouse and rat heart. *Nat. Protoc.* **9**, 1662–1681 (2014).

BRIEF COMMUNICATIONS ARISING

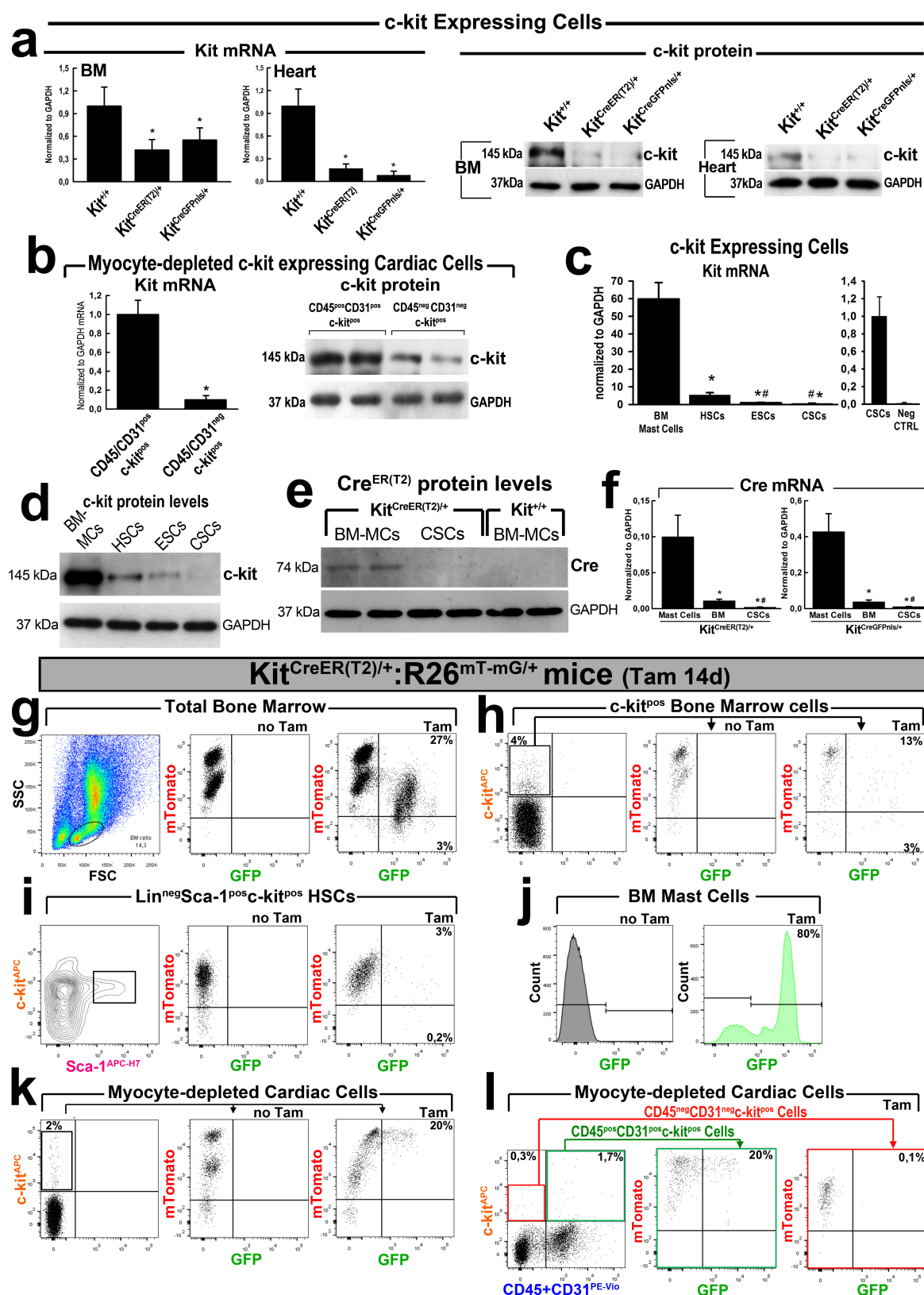
13. Vicinanza, C. *et al.* Adult cardiac stem cells are multipotent and robustly myogenic: c-kit expression is necessary but not sufficient for their identification. *Cell Death Differ.* **24**, 2101–2116 (2017).
14. Muzumdar, M. D., Tasic, B., Miyamichi, K., Li, L. & Luo, L. A global double-fluorescent Cre reporter mouse. *Genesis* **45**, 593–605 (2007).
15. Bernex, F. *et al.* Spatial and temporal patterns of c-kit-expressing cells in WlacZ/+ and WlacZ/WlacZ mouse embryos. *Development* **122**, 3023–3033 (1996).

Supplementary Information accompanies this Comment.

Author Contributions D.T. and B.N.-G. designed the research studies, analysed data and wrote the manuscript. C.V., I.A., E.C., M.S., F.M., T.M., F.F., E.D.G., E.I., P.M., V.A. and K.U. conducted the experiments, acquired and analysed data. F.C., A.T. and P.V. conducted RNA-sequencing experiments and analysed bioinformatic data. R.L., A.M.I., D.S. and C.I. contributed reagents. All authors approved the manuscript.

Competing Interests Declared none.

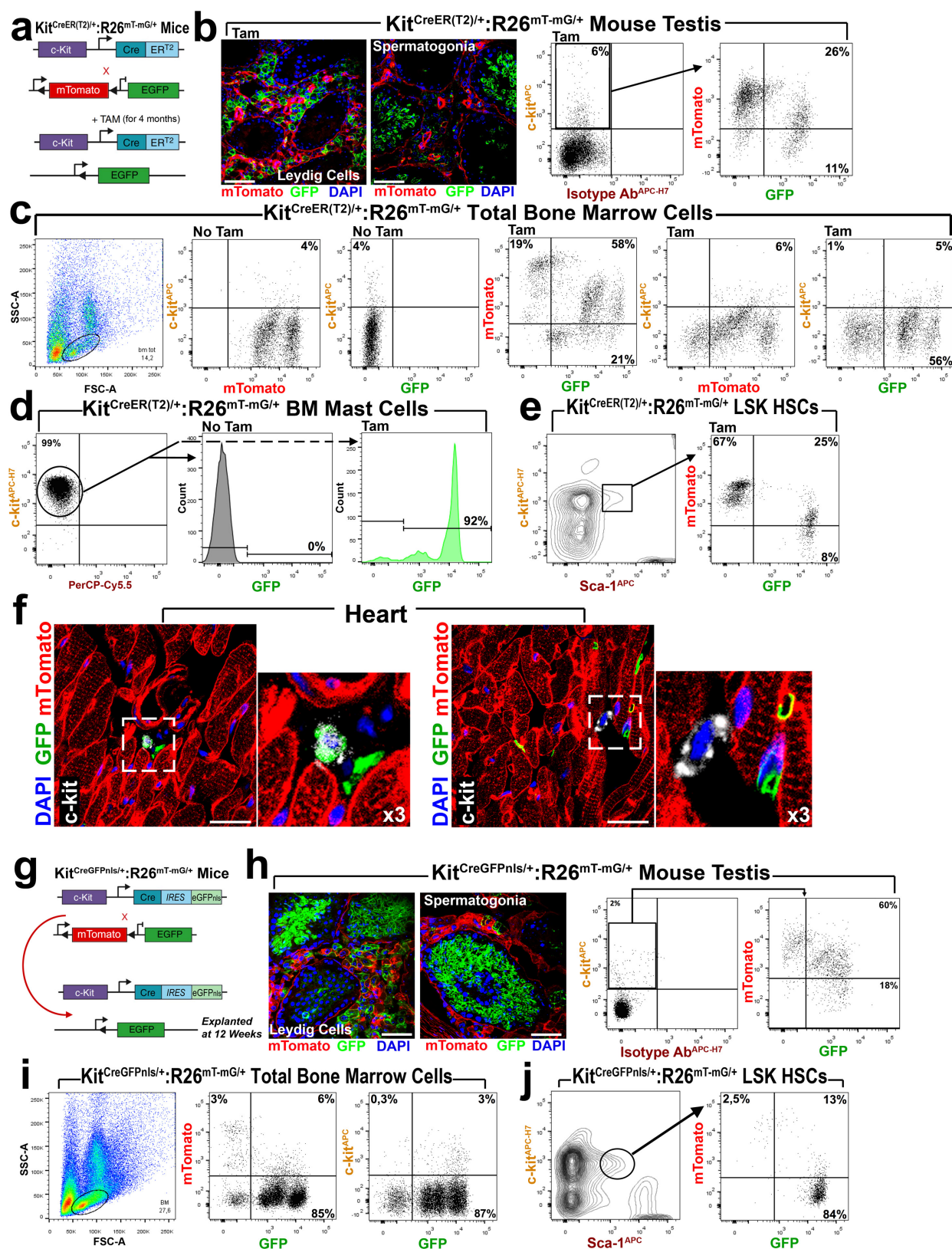
doi:10.1038/nature25771



Extended Data Figure 1 | See next page for caption.

Extended Data Figure 1 | The *Kit^{CreER(T2)}* allele reduces c-Kit expression and does not fate-map c-Kit-expressing CSCs with a short-term tamoxifen diet. **a**, *Kit* mRNA (left) and c-Kit protein (right) levels in all c-Kit-expressing cells isolated from bone marrow (BM) and heart of wild-type *Kit^{+/+}*, heterozygous *Kit^{CreER(T2)/+}* and *Kit^{CreGFPnl/+}* mice. Representative of $n = 5$ bone marrow samples and hearts. $*P < 0.05$ versus wild-type *Kit^{+/+}* mice. **b**, qPCR (left) cumulative data and representative western blot (right) analysis show that *Kit* mRNA and c-Kit protein levels in cardiomyocyte-depleted FACS-sorted $CD45^{pos}CD31^{pos}$ -*Kit^{pos}* cardiac cells versus $CD45^{neg}CD31^{neg}$ -*Kit^{pos}* cardiac cells. Representative of $n = 5$ hearts. $*P < 0.05$ versus $CD45^{pos}CD31^{pos}$ -*Kit^{pos}* cardiac cells. **c, d**, *Kit* mRNA (**c**) and c-Kit protein (**d**) expression levels in different cell populations from wild-type mice. Representative of $n = 4$ biological replicates. $*P < 0.05$ versus mast cells; $\#P < 0.05$ versus HSCs in **c**. **c**, *Kit* transcript levels are verified over *Kit* mRNA absence in negative control (Neg CTRL), which are triple-negative-sorted $CD45^{neg}CD31^{neg}$ -*Kit^{neg}* cardiac cells. **e**, Western blot analysis showing *Cre^{ER(T2)}* expression in bone marrow mast cells; *Cre^{ER(T2)}* expression is only faintly detectable in freshly isolated $CD45^{neg}CD31^{neg}$ -*Kit^{pos}* CSCs from *Kit^{CreER(T2)/+}* mice. Representative of $n = 3$ biological replicates. **f**, *cre* mRNA levels in bone marrow mast cells, bone marrow *Lin^{neg}* cells and CSCs from *Kit^{CreER(T2)/+}*

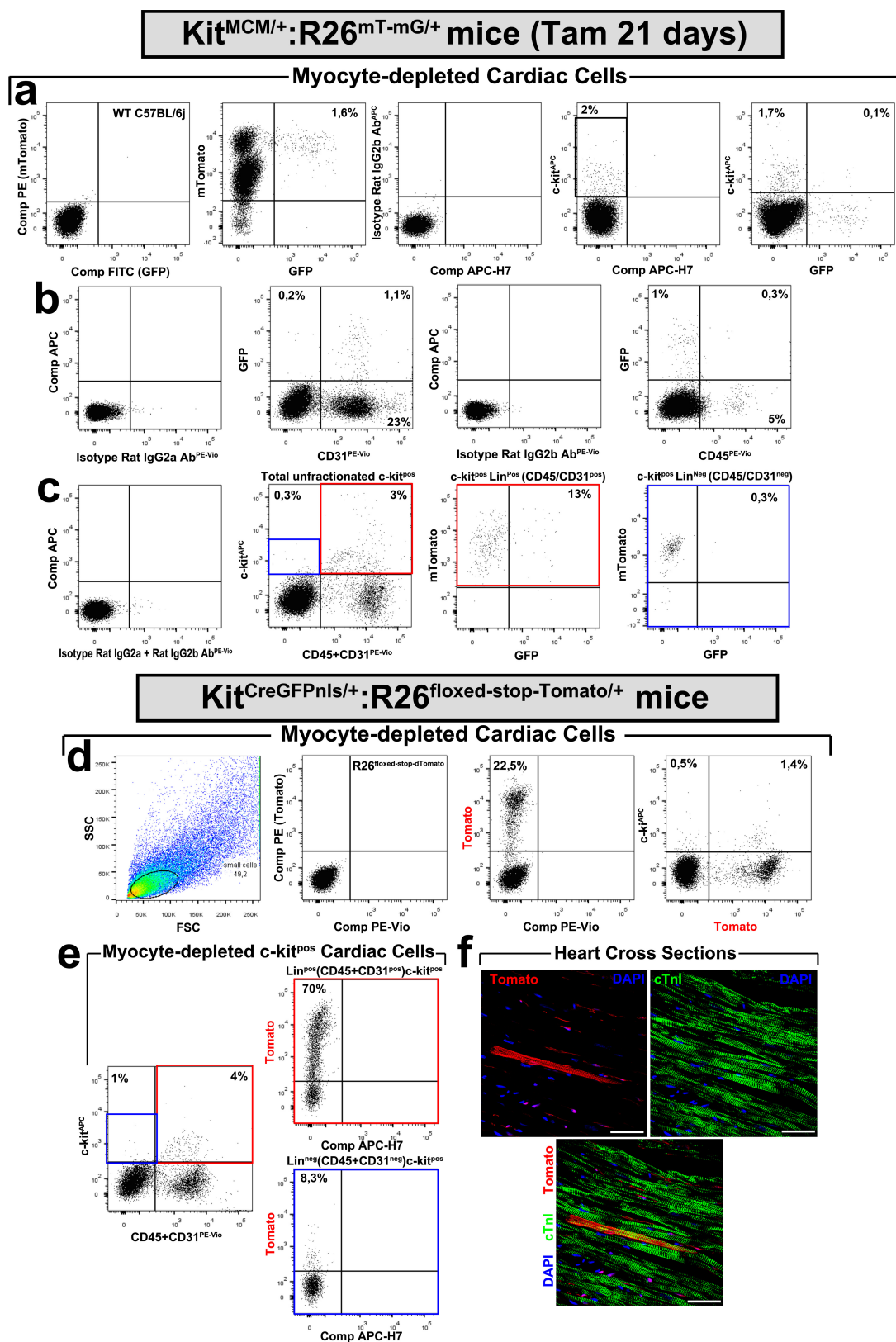
and *Kit^{CreGFPnl/+}* mice. Representative of $n = 5$ mice. $*P < 0.05$ versus mast cells; $\#P < 0.05$ versus bone marrow. **g–l**, FACS analysis showing recombination efficiency in isolated bone marrow and cardiac cells from double-mutant *Kit^{CreER(T2)/+};Rosa26^{mT/mG/+}* mice treated with standard tamoxifen (Tam) diet (400 mg per kg diet) for 14 days. **g**, Fraction of Cre-recombined GFP-expressing cells in the monocyte–lymphocyte gate of total bone marrow cells after 14 days of tamoxifen diet. No recombination is evidence in the absence of tamoxifen. **h**, Fraction of Cre-recombined GFP-expressing cells in c-*Kit^{pos}* bone marrow cells after 14 days of tamoxifen diet. **i**, Gating of the *Lin^{neg}* bone marrow cells for the long-term repopulating *Lin^{neg}Sca-1^{pos}*-*Kit^{pos}* (LSK) HSCs showing that only $<5\%$ HSCs showed recombination. **j**, The majority of bone marrow mast cells showed recombination and expressed GFP after 14 days of tamoxifen diet. **k**, The 14-day tamoxifen regime recombined $\leq 20\%$ of the cardiomyocyte-depleted total c-*Kit^{pos}* cardiac cell population. **l**, All GFP-recombined c-*Kit^{pos}* cardiac cells were positive for CD45, CD31 ($CD45^{pos}$, $CD31^{pos}$) or both, representing cardiac mast cells or endothelial (progenitor) cells. Only a negligible fraction, $<1\%$, of the c-*Kit^{low}* and CSC-enriched *Lin^{neg}CD45^{neg}CD31^{neg}*-*Kit^{pos}* cardiac cells showed recombination and expression of GFP. **g–l**, Representative of $n = 5$ bone marrow samples and hearts. Data are mean \pm s.d.



Extended Data Figure 2 | See next page for caption.

Extended Data Figure 2 | Fate-map tracking of c-Kit^{pos} cells using *Kit^{cre}* mice. **a**, Mice with KI in exon 1 of the *Kit* locus targeted to express tamoxifen-inducible Cre recombinase (CreER^(T2)) were crossed with Cre-reporter Rosa26 mice (*Rosa26^{mT/mG}*); the double-heterozygous mice were fed tamoxifen for four months. **b**, Confocal images (left) and FACS plots (right) show that *Kit^{CreER(T2)}* allele recombines in c-Kit-expressing Leydig cells and spermatogonia (and their derivatives within the seminiferous tubule) to express GFP in the testis whereas GFP expression was induced in around 40% of c-Kit^{pos} cells of this tissue. Representative of *n* = 5 mice. Scale bar, 50 μm. **c**, FACS plot of total bone marrow (with lymphocyte–monocyte gating strategy, left) of double-mutant *Kit^{CreER(T2)/+};*Rosa26^{mT/mG}/+** showing GFP expression in approximately 80% of cells, but no signal in the absence of tamoxifen. In particular, ≥70% of all c-Kit-antibody-detected cells were GFP^{pos}. Representative of *n* = 5 bone marrow samples. **d**, Most bone marrow mast cells (c-Kit expression, left) showed recombination and expressed GFP. **e**, Only a fraction (<35%) of the long-term LSK HSCs expressed GFP. **f**, Representative confocal images of cardiac cross-sections

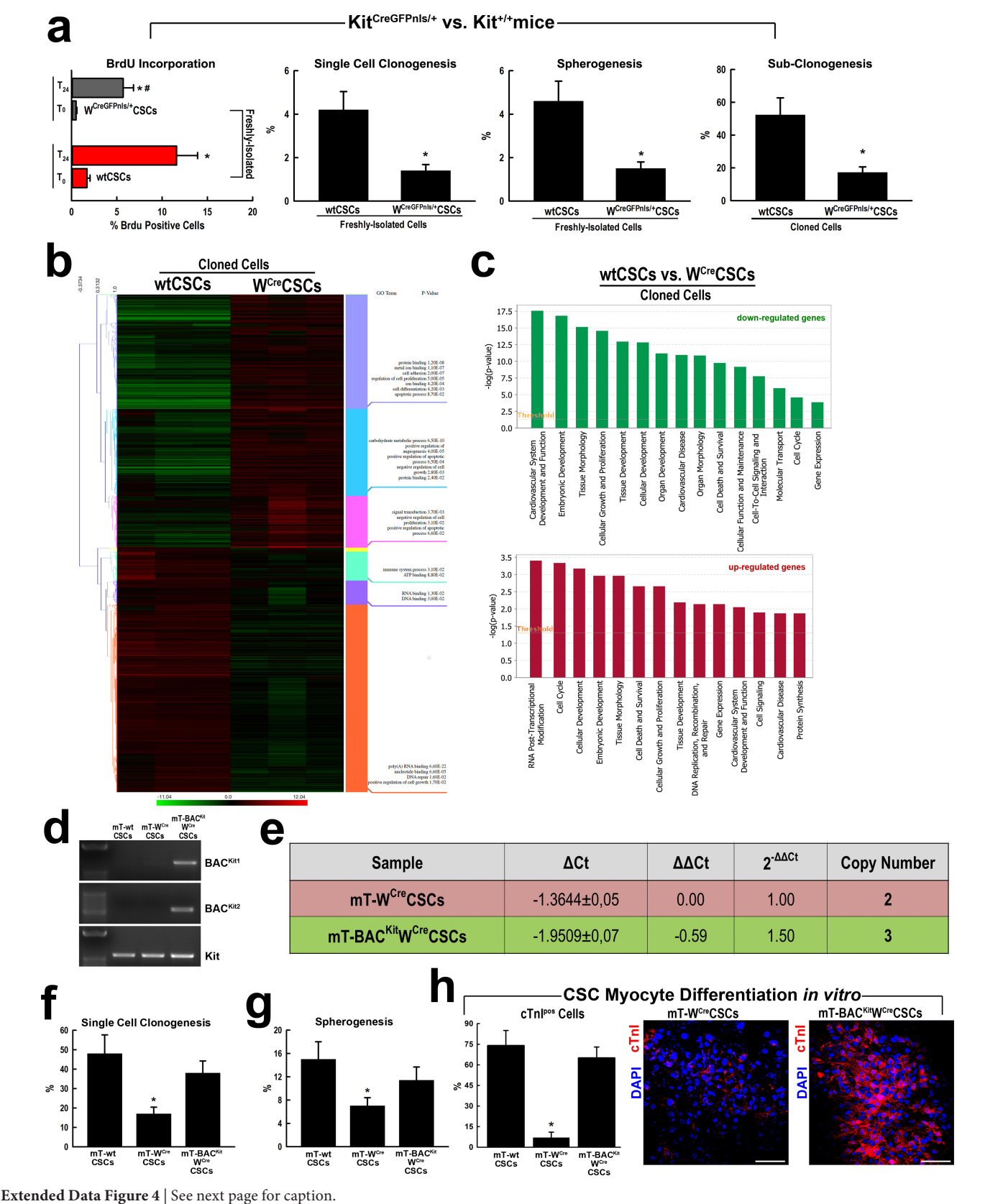
from *Kit^{CreER(T2)/+};*Rosa26^{mT/mG}/+** mice treated with tamoxifen diet for four months show that Cre-dependent recombination-induced GFP expression is detectable only in a fraction (around 60%) of non-cardiomyocyte c-Kit-labelled cardiac cells. Scale bars, 20 μm. Representative of *n* = 5 mice. **g**, Mice with *Kit* exon 1 locus KI to express Cre recombinase and GFP with a nuclear localization sequence (GFPnls) behind an internal ribosome entry site (IRES) were crossed to Cre reporter Rosa26 mice (*Rosa26^{mT/mG}*) for lineage tracing. **h**, Confocal images (left) and FACS plots (right) show that the *Kit^{CreGFPnls}* allele recombines in Leydig cells and spermatogonia in the testis to become GFP^{pos} whereas the overall recombination in the testis was approximately 80% of the c-Kit^{pos} cells. Scale bars, 50 μm. Note that the majority of recombined cells still express mT. Representative of *n* = 5 mice. **i**, FACS plot of total bone marrow cells from double-mutant *Kit^{CreGFPnls/+};*Rosa26^{mT/mG}/+** mice showing GFP expression in more than 90% of cells. Approximately 90% of total c-Kit-antibody-detected cells were GFP^{pos}. **j**, A majority of LSK HSCs expressed GFP. Representative of *n* = 5 bone marrow samples.



Extended Data Figure 3 | See next page for caption.

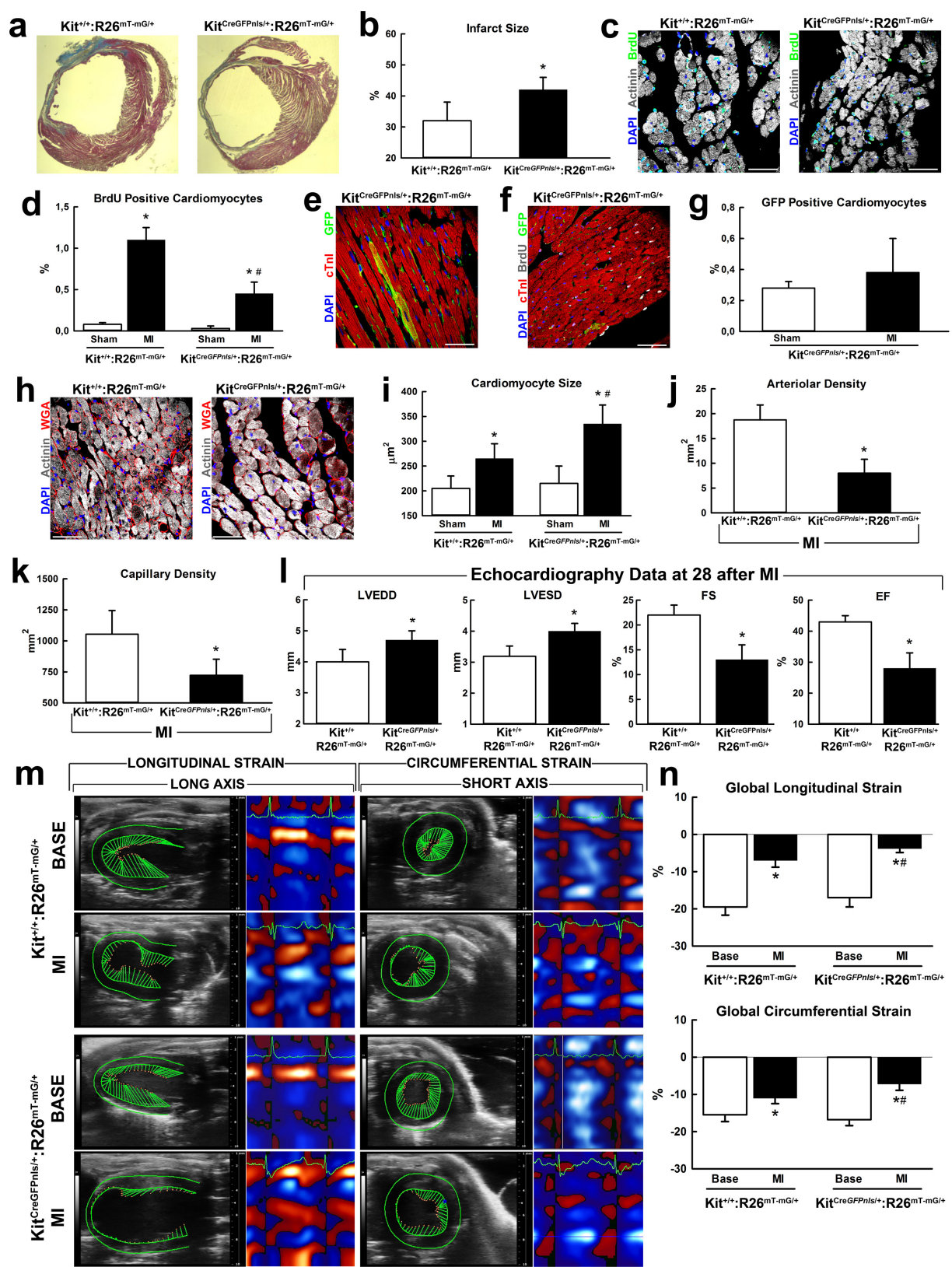
Extended Data Figure 3 | Fate-map tracking of c-Kit^{pos} CSCs using *Kit^{cre}* mice. FACS analysis showing recombination efficiency in isolated cardiomyocyte-depleted cardiac cells from double-mutant *Kit^{MCM/+};Rosa26^{mT/mG/+}* mice treated with a standard tamoxifen diet (400 mg per kg diet) for 21 days ($n = 5$ mice). **a**, After 21 days of tamoxifen diet, less than 2% of total cardiac cells showed recombination and induction of GFP. Note that all the identified MCM-recombined GFP^{pos} cells were also still mT^{pos}. Approximately 10% of the cardiomyocyte-depleted c-Kit^{pos} cardiac cell population showed recombination and expressed GFP. **b**, Independently of the level of c-Kit expression, almost all recombined GFP^{pos} cells were CD31^{pos} (around 85%) and CD45^{pos} (about 25%). Thus, a standard tamoxifen regime with a widely used pulse period only genetically recombines lineage-committed cells (endothelial or immune cells) within the heart. **c**, All GFP-expressing recombined c-Kit^{pos} cardiac cells were CD45^{pos}, CD31^{pos} or both, representing cardiac mast cells or endothelial (progenitor) cells. Only a negligible fraction, <1%, of the c-Kit^{low} and CSC-enriched

Lin^{neg}CD45^{neg}CD31^{neg}c-Kit^{pos} cardiac cells showed recombination and induction of GFP expression. **a–c**, Representative of $n = 4$ hearts. **d–f**, *Kit^{CreGFPnls/+}* mice were crossed with B6;129S6-Gt(ROSA)26Sortm9(CAG-tdTomato)Hze/J (abbreviated as *Rosa26^{floxex-stop-tdTomato}*) mice that have a targeted mutation in the Gt(ROSA)26Sor locus with a *loxP*-flanked STOP cassette preventing transcription of a CAG-promoter-driven red fluorescent protein variant (tdTomato), which is expressed following Cre-mediated recombination. **d**, Around 20% of total cardiac cells and about 70% of c-Kit^{pos} cardiac cells from these mice are tdTomato^{pos}. **e**, Around 70% of lineage-committed endothelial/mast cell (CD45^{pos}CD31^{pos}c-Kit^{pos}) cardiac cells showed recombination and expressed tdTomato, whereas less than 10% of the CSC-enriched CD45^{neg}CD31^{neg}c-Kit^{pos} showed recombination and were tdTomato^{pos}. **d, e**, Representative of $n = 5$ hearts. **f**, Confocal image showing recombined tdTomato^{pos} cardiomyocytes in the ventricular myocardium of 8–12-week-old *Kit^{CreGFPnls/+};Rosa26^{floxex-stop-tdTomato/+}* mice. Representative of $n = 3$ hearts. Scale bar, 50 μ m.



Extended Data Figure 4 | W^{cre} allele KI and resulting *Kit* heterozygosity impair CSC biology and myogenic potential *in vitro*. **a**, Serum-induced 24 h BrdU incorporation by freshly isolated $W^{CreGFPnl/+}$ CSCs compared to wild-type CSCs. $*P < 0.05$ versus T_0 ; $\#P < 0.05$ versus wild-type CSCs. Clonal efficiency of freshly isolated $W^{CreGFPnl/+}$ CSCs compared to wild-type CSCs. $*P < 0.05$ versus wild-type CSCs. $W^{CreGFPnl/+}$ CSCs formed fewer cardiospheres than wild-type CSCs. The number of cardiospheres is expressed as a percentage of 1×10^5 plated cells. $*P < 0.05$ versus wild-type CSCs. Single-cell-cloned $W^{CreGFPnl/+}$ CSCs re-cloned at lower efficiency than wild-type CSCs. $*P < 0.05$ versus wild-type CSCs. $n = 5$ experiments. **b**, Heat map of RNA-sequencing profile of the 2,425 downregulated and 2,870 upregulated genes for the comparison of W^{Cre} and wild-type CSCs and their gene ontology clustering for specific gene function. $n = 3$ of biological replicates. **c**, Functional categorization by ingenuity pathway analysis (IPA) of the downregulated and upregulated genes by RNA sequencing in the comparison between W^{Cre} CSC (from $Kit^{CreER(T2)/+}$ mice) and wild-type CSC clones. Histogram represents the

most significant canonical pathways generated using IPA software. The ratio was calculated by dividing the number of genes from our dataset that map to each single pathway by the total number of genes included into the canonical pathway. **b, c**, Mean of $n = 3$ biological replicates. **d**, PCR analysis with two different pair of BAC primers showing BAC^{Kit} expression in mT- W^{Cre} CSCs. *Kit* gene primers were used as control. $n = 3$ biological replicates. **e**, qPCR data showing *Kit* DNA copy number in BAC-naive mT- W^{Cre} CSCs and mT-BAC^{Kit} W^{Cre} CSCs. Representative of $n = 3$ biological replicates. **f**, Clonal efficiency of mT-BAC^{Kit} W^{Cre} CSCs, mT- W^{Cre} CSCs and mT-WT CSCs. $*P < 0.05$ versus all other treatments. $n = 5$ biological replicates. **g**, Cardiosphere formation of mT-BAC^{Kit} W^{Cre} CSCs, mT- W^{Cre} CSCs and mT-WT CSCs. $*P < 0.05$ versus all other treatments. $n = 5$ biological replicates. **h**, Bar graph and confocal images of cTnI expression in cardiospheres from cloned mT-WT CSCs, mT- W^{Cre} CSCs and mT-BAC^{Kit} W^{Cre} CSCs after 14 days in myogenic medium. $*P < 0.05$ versus all other treatments. Representative of $n = 5$ biological replicates. Scale bar, 50 μ m. Data are mean \pm s.d.

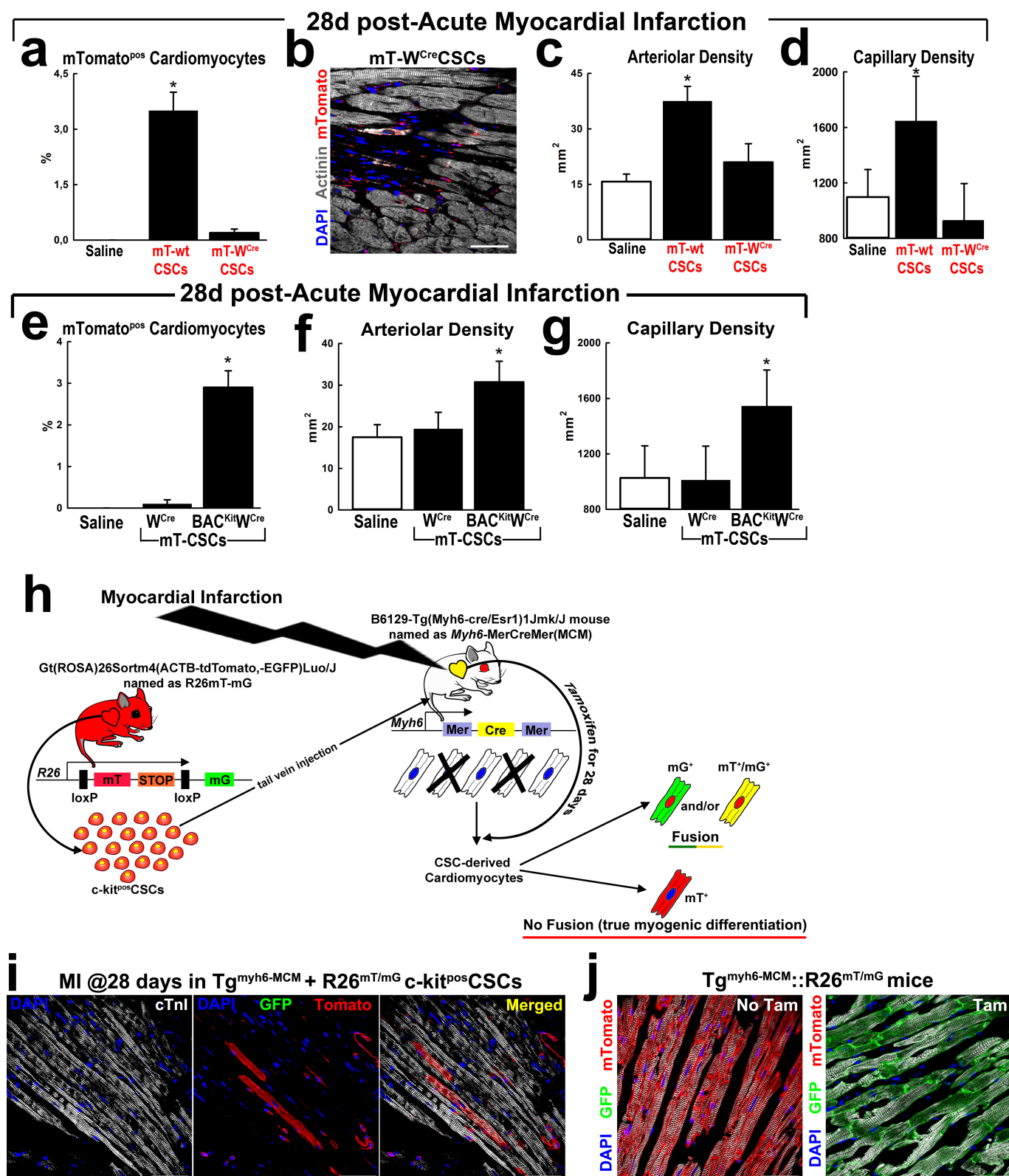


Extended Data Figure 5 | See next page for caption.

Extended Data Figure 5 | *Kit^{cre}*-KI hampers cardiac regeneration and repair after myocardial infarction.

To test the effect of the *Kit^{cre}*-KI mutations on cardiac regeneration *in vivo*, 10-week-old female *Kit^{CreGFPnls/+};*Rosa26^{mT/mG}/+** double-heterozygous mice and *Rosa26^{mT/mG}/+* heterozygous controls were subjected to myocardial infarction (MI) by permanent ligation of the left coronary artery followed by systemic BrdU administration through mini-osmotic pumps. *Kit^{CreGFPnls/+}*-KI mice were used instead of tamoxifen-inducible *Kit^{CreER(T2)/+}*-KI mice, because of their highest *Kit^{cre}*-induced recombination and to avoid possible confounding effects related to tamoxifen (and tamoxifen-induced Cre) toxicity *in vivo*⁹. *Kit^{CreGFPnls/+};*Rosa26^{mT/mG}/+** double-heterozygous females had to be used, because in our hands the acute mortality of myocardial infarction in sibling males was $\geq 75\%$ whereas it was around 30% in the *Rosa26^{mT/mG}/+* heterozygous controls. Additionally, around 50% of *Kit^{CreGFPnls/+};*Rosa26^{mT/mG}/+** males and about 30% of *Kit^{CreGFPnls/+};*Rosa26^{mT/mG}/+** females showed spontaneous alterations of cardiac function at baseline before myocardial infarction and had to be excluded. **a**, Light microscopy images showing Masson staining of infarcted hearts from c-Kit wild-type control *Kit^{+/+};*Rosa26^{mT/mG}/+** and *Kit^{CreGFPnls/+};*Rosa26^{mT/mG}/+** mice. Representative of $n = 6$ mice per group. **b**, Infarct size assessment 28 days after coronary ligation in *Kit^{+/+};*Rosa26^{mT/mG}/+** and *Kit^{CreGFPnls/+};*Rosa26^{mT/mG}/+** mice. $*P < 0.05$ versus *Kit^{+/+};*Rosa26^{mT/mG}/+** mice. $n = 6$ mice per group. **c**, Representative confocal microscopy images of BrdU incorporation in the border zone of infarcted hearts from *Kit^{+/+};*Rosa26^{mT/mG}/+** and *Kit^{CreGFPnls/+};*Rosa26^{mT/mG}/+** mice. Scale bar, 50 μm . **d**, Number of newly generated BrdU^{pos} cardiomyocytes in sham and infarcted (MI) *Kit^{+/+};*Rosa26^{mT/mG}/+** and *Kit^{CreGFPnls/+};*Rosa26^{mT/mG}/+** mice 28 days after sham or myocardial infarction surgery. $*P < 0.05$ versus sham; $\#P < 0.05$ versus *Kit^{+/+};*Rosa26^{mT/mG}/+** mice. $n = 6$ mice per group. **e, f**, Representative confocal microscopy images of GFP^{pos} cardiomyocytes

in the border zone of infarcted hearts from *Kit^{CreGFPnls/+};*Rosa26^{mT/mG}/+** mice. Note that in **f** GFP^{pos} cardiomyocytes are BrdU^{neg}. Scale bar, 50 μm . **g**, Bar graph with cumulative data showing the number of GFP^{pos} cardiomyocytes in sham and infarcted *Kit^{CreGFPnls/+};*Rosa26^{mT/mG}/+** mice 28 days after surgery. **h**, Representative confocal images of cardiac cross-sections showing higher cardiomyocyte hypertrophy in *Kit^{CreGFPnls/+};*Rosa26^{mT/mG}/+** compared to *Kit^{+/+};*Rosa26^{mT/mG}/+** mice 28 days after myocardial infarction. WGA, wheat-germ agglutinin. Scale bar, 50 μm . **i**, Cardiomyocyte size in *Kit^{+/+};*Rosa26^{mT/mG}/+** and *Kit^{CreGFPnls/+};*Rosa26^{mT/mG}/+** mice. $*P < 0.05$ versus sham; $\#P < 0.05$ versus *Kit^{+/+};*Rosa26^{mT/mG}/+** mice. $n = 6$ mice per group. **j**, Arteriolar density in the infarct border zone of *Kit^{+/+};*Rosa26^{mT/mG}/+** and *Kit^{CreGFPnls/+};*Rosa26^{mT/mG}/+** mice 28 days after myocardial infarction. $*P < 0.05$ versus *Kit^{+/+};*Rosa26^{mT/mG}/+** mice. $n = 6$ mice per group. **k**, Capillary density in the infarct border zone of *Kit^{+/+};*Rosa26^{mT/mG}/+** and *Kit^{CreGFPnls/+};*Rosa26^{mT/mG}/+** mice 28 days after myocardial infarction. $*P < 0.05$ versus *Kit^{+/+};*Rosa26^{mT/mG}/+** mice. $n = 6$ mice per group. **l**, Echocardiography assessment of left ventricular function 28 days after myocardial infarction in *Kit^{+/+};*Rosa26^{mT/mG}/+** and *Kit^{CreGFPnls/+};*Rosa26^{mT/mG}/+** mice. $*P < 0.05$ versus *Kit^{+/+};*Rosa26^{mT/mG}/+** mice. $n = 6$ mice per group. LVEDD, left ventricle end diastolic diameter; LVESD, left ventricle end systolic diameter; EF, ejection fraction; FS, fractional shortening. **m**, Representative echo images and longitudinal and circumferential strain traces in long and short axis, from *Kit^{+/+};*Rosa26^{mT/mG}/+** and *Kit^{CreGFPnls/+};*Rosa26^{mT/mG}/+** mice at baseline (Base) and 28 days after myocardial infarction. **n**, Longitudinal and circumferential strain values in *Kit^{+/+};*Rosa26^{mT/mG}/+** versus *Kit^{CreGFPnls/+};*Rosa26^{mT/mG}/+** mice at baseline and 28 days after myocardial. $*P < 0.05$ versus baseline; $\#P < 0.05$ versus *Kit^{+/+};*Rosa26^{mT/mG}/+** mice. $n = 6$ mice per group. Data are mean \pm s.d.



Extended Data Figure 6 | See next page for caption.

Extended Data Figure 6 | W^{cre} -null allele hampers CSC regenerative potential after myocardial infarction and neomyogenesis, which is independent from cell fusion. **a**, Cumulative data of CSC-derived mT^{pos} cardiomyocytes from female C57BL/6J wild-type mice 28 days after myocardial infarction treated with either saline ($n = 6$ mice) or with the indicated CSC types ($n = 6$ mice per group). $*P < 0.05$ versus all other treatments. **b**, Representative confocal image of a CSC-derived newly formed mT^{pos} cardiomyocyte in mT- W^{cre} CSC-injected mice 28 days after myocardial infarction. Scale bar, 30 μm . **c**, Cumulative data of arteriolar density in the infarct border zone of saline-, mT-WT CSC- or mT- W^{cre} CSC-treated female C57BL/6J wild-type mice 28 days after myocardial infarction. $*P < 0.05$ versus all other treatments. **d**, Cumulative data of capillary density in the infarct border zone of saline-, mT-WT CSC- or mT- W^{cre} CSC-treated female C57BL/6J wild-type mice 28 days after myocardial infarction. $*P < 0.05$ versus all other treatments. **b–d**, Number of mice per group as indicated in **a**. **e–g**, Cumulative data of CSC-derived mT^{pos} cardiomyocytes (**e**), arteriolar density (**f**) and capillary density (**g**) in the infarct border zone of female wild-type C57BL/6J mice 28 days after myocardial infarction treated with saline ($n = 6$), mT- W^{cre} CSCs ($n = 6$) or mT-BAC^{Kit} W^{cre} CSCs ($n = 6$). $*P < 0.05$ versus all other treatments. **h**, Schematic of study design to assess CSC fusion with hosting cardiomyocytes. Transgenic (Tg) *Myh6-MerCreMer* ($Tg^{Myh6-MCM}$) female ($n = 3$) mice that carry a transgenic cardiomyocyte-restricted tamoxifen-inducible *cre* gene construct, had were subjected to myocardial infarction by coronary ligation. Right after coronary ligation, mice were treated by tail-vein injection with wild-type clonal mT^{pos} CSCs (mT-WT CSCs). Mice were fed with a tamoxifen diet for four weeks. The mT-WT CSCs were isolated from *Rosa26^{mT/mG}* male mice and cloned. Therefore, these clonal cells and their progeny constitutively express membrane Tomato (mT^{pos}), which switches to express membrane GFP (becoming mG^{pos}) when recombined in response to Cre recombinase. The injection of mT-WT

CSCs in $Tg^{Myh6-MCM}$ mice tests directly whether new cardiomyocytes are exclusively the progeny of the mT-WT CSCs injected, in which case these cells should be red (mT^{pos}), or the result of cell fusion of the injected cells with host cardiomyocytes, in which case the cells should be yellow (mT^{pos} and mG^{pos} together show as yellow) or only green (mG^{pos}). Indeed, if the putative new cardiomyocytes were not newly generated, but the product of the fusion of the injected cells with the host cardiomyocytes, these fused cells should be yellow or green. Tamoxifen activates Cre only in the host cardiomyocytes of $Tg^{Myh6-MCM}$ mice and the recombination induces the expression of mG only if the injected reporter-switchable CSCs fuse with the host cardiomyocytes. Thus, the recombined host cardiomyocytes will be either mT^{pos} and mG^{pos} (yellow) or mG^{pos} (green), the latter depending on the dilution time of mT expression after recombination. On the other hand, mG^{neg}mT^{pos} cardiomyocytes can only be the direct progeny of the injected CSCs with no fusion to host cardiomyocytes. **i**, Representative confocal microscopy images of CSC-derived mT^{pos} cardiomyocytes in the border zone ($3.0 \pm 0.5\%$) of $Tg^{Myh6-MCM}$ mice with myocardial infarction treated with mT-WT CSCs and fed with a tamoxifen diet. It is evident that labelled cardiomyocytes in the infarct border zone were mT^{pos} but mG^{neg} as expected for the cardiomyocyte-differentiated progeny of the injected CSCs, which excludes cell fusion as a relevant mechanism for the appearance of these new muscle cells after myocardial infarction. Scale bar, 50 μm . Note, the level of baseline red or green autofluorescent cardiomyocytes in un-transplanted mice resulting from the staining protocols was $<0.005\%$. None of these weak apparently positive signals had the typical membrane localization expected for the membrane-targeted dTomato and GFP Cre reporters. **j**, The mT and mG signal gains for the confocal images in **i** were set to appropriate positive controls using double transgenic $Tg^{Myh6-MCM}::Rosa26^{mT/mG}$ mice fed with a normal diet (no tamoxifen) or tamoxifen diet. Scale bar, 50 μm . Data are mean \pm s.d.

van Berlo *et al.* replyREPLYING TO: C. Vicinanza *et al.* *Nature* **555** <https://doi.org/10.1038/nature25771> (2018)

The accompanying Comment by Vicinanza *et al.*¹ does not agree with our previous study² regarding the utility of c-Kit^{pos} (also known as Kit^{pos}) cells when investigating regeneration in the heart. The authors claim that three previous publications^{2–4} using *Kit^{cre}* knock-in lineage-tracing question their previous work that shows that tissue-specific c-Kit^{pos} cardiac stem/progenitor cells (CSCs) are necessary and sufficient for cardiomyocyte regeneration and/or replenishment after injury⁵. However, the regenerative data within their previous manuscript did not address the role of endogenous c-Kit^{pos} cells in the heart⁵. Instead, they relied on clonally derived c-Kit^{pos} cells that were selected in culture and then injected into the circulatory system of mice and these cells remarkably homed to the heart and generated abundant new cardiomyocytes that repaired isoproterenol-mediated injury⁵.

The authors now claim that the number of c-Kit^{pos} cells with cardiomyogenic-clone-producing potential is very low in the mouse heart (approximately 1% of total c-Kit^{pos} cells)⁶. Indeed, the authors showed that broadly isolated primary cardiac c-Kit^{pos} cells, that were not clonally selected, had almost no ability to generate new cardiomyocytes or repair the infarcted heart⁶. The authors' new interpretation¹ that only a very small portion of c-Kit^{low} cells can be made clonal with regenerative activity contradicts their published results from 2003 in which broadly isolated c-Kit^{pos} cells regenerated 70% of the myocardium after infarction injury⁷, or published results from 2013 showing that the majority of endogenous c-Kit^{pos} cells express Nkx2.5 (a cardiomyocyte-determining transcription factor)⁵.

The authors further claim that *Kit^{cre}* knock-in lineage tracing approaches^{2–4} fail to track these rare c-Kit^{low} CSCs, because of poor recombination efficiency and heterozygosity in the *Kit* allele. While we have previously addressed this criticism and discussed data to the contrary⁸, we have recently published additional data showing that the majority of c-Kit^{pos}CD45^{neg} cells isolated from the hearts of *Kit^{cre}* mice are correctly lineage-traced⁹, suggesting that this technique for tracking cells in mice is effective^{2–4}. Indeed, a working group of established cardiac scientists recently published a consensus statement confirming that endogenous cardiac c-Kit^{pos} cells are an unlikely source for meaningful heart regeneration¹⁰. This statement was also based on the known biology of the adult mammalian heart, which essentially lacks acute cardiomyocyte-forming regenerative capacity. However, it seems reasonable that one could select and isolate rare clonally derived

c-Kit^{pos} progenitor cell lines with ectopic cardiomyogenic activity, as suggested in the Comment¹. However, the Comment does not address the role of endogenous c-Kit^{pos} cells in the heart. The current negative results simply address whether clonally derived c-Kit-expressing cells can have cardiomyogenic capacity, which we believe lacks all *in vivo* relevance and in no way contests our previous observations.

Jop H. van Berlo¹, Onur Kanisicak², Marjorie Maillet²,
Ronald J. Vagnozzi², Jason Karch², Suh-Chin J. Lin²,
Ryan C. Middleton³, Eduardo Marbán³ & Jeffery D. Molkentin^{2,4}

¹Department of Medicine, Division of Cardiology, Lillehei Heart Institute, University of Minnesota, Minneapolis, Minnesota, USA.

²Department of Pediatrics, Cincinnati Children's Hospital Medical Center, Cincinnati, Ohio, USA.

email: Jeff.Molkentin@cchmc.org

³Cedars-Sinai Heart Institute, 8700 Beverly Boulevard, Los Angeles, California, USA.

⁴Howard Hughes Medical Institute, Cincinnati Children's Hospital Medical Center, Cincinnati, Ohio, USA.

1. Vicinanza, C. *et al.* *Kit^{cre}* knock-ins fail to fate-map cardiac stem cells. *Nature* **555**, <https://doi.org/10.1038/nature25771> (2018).
2. van Berlo, J. H. *et al.* c-kit⁺ cells minimally contribute cardiomyocytes to the heart. *Nature* **509**, 337–341 (2014).
3. Sultana, N. *et al.* Resident c-kit⁺ cells in the heart are not cardiac stem cells. *Nat. Commun.* **6**, 8701 (2015).
4. Liu, Q. *et al.* Genetic lineage tracing identifies *in situ* Kit-expressing cardiomyocytes. *Cell Res.* **26**, 119–130 (2016).
5. Ellison, G. M. *et al.* Adult c-kit^{pos} cardiac stem cells are necessary and sufficient for functional cardiac regeneration and repair. *Cell* **154**, 827–842 (2013).
6. Vicinanza, C. *et al.* Adult cardiac stem cells are multipotent and robustly myogenic: c-kit expression is necessary but not sufficient for their identification. *Cell Death Differ.* **24**, 2101–2116 (2017).
7. Beltrami, A. P. *et al.* Adult cardiac stem cells are multipotent and support myocardial regeneration. *Cell* **114**, 763–776 (2003).
8. Molkentin, J. D. Letter by Molkentin regarding article, "The absence of evidence is not evidence of absence: the pitfalls of Cre knock-ins in the c-Kit locus". *Circ. Res.* **115**, e21–e23 (2014).
9. Kanisicak, O., Vagnozzi, R. J. & Molkentin, J. D. Identity crisis for regenerative cardiac cKit⁺ cells. *Circ. Res.* **121**, 1130–1132 (2017).
10. Eschenhagen, T. *et al.* Cardiomyocyte regeneration: a consensus statement. *Circulation* **136**, 680–686 (2017).

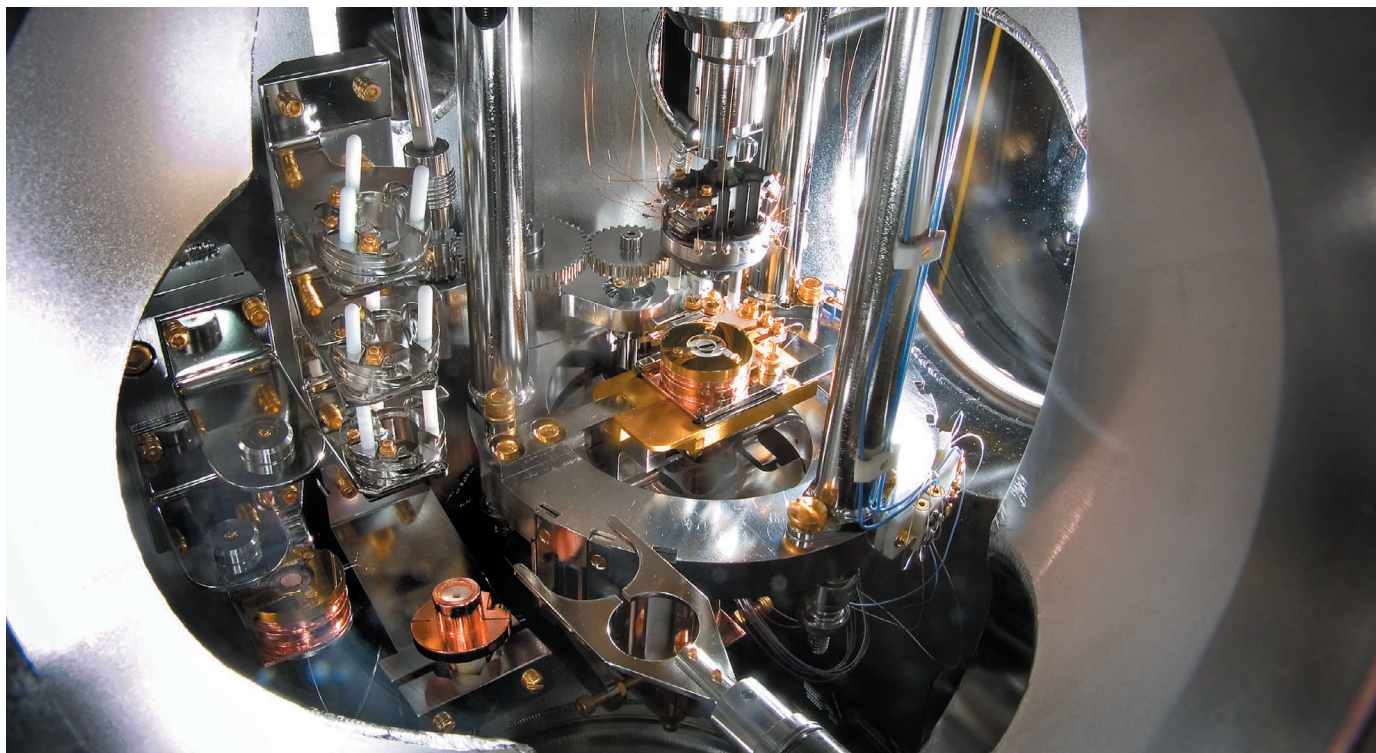
doi:10.1038/nature25772

TECHNOLOGY FEATURE

THE ATOMIC-FORCE REVOLUTION

Atomic force microscopy is revealing molecular structures with startling clarity. Artificial intelligence and automation could expand its potential.

BROOKHAVEN NATIONAL LABORATORY



The atomic force microscope enables researchers to image the surfaces of materials at atomic resolution.

BY ANDY EXTANCE

While collecting images of the molecular structure of an antiseptic compound called 8-hydroxyquinoline, nanoscientist Xiaohui Qiu saw something that he didn't expect.

His team was using a method called atomic force microscopy, dragging an ultrasharp probe across the surface of a sample to reveal its topography at subnanometre resolution. The technique had previously revealed single molecules, complex mixtures and even chemical reactions in extraordinary detail. But Qiu's image of 8-hydroxyquinoline contained something peculiar: connecting silhouettes of the organic molecule's two rings were fine tendrils of contrast in regions where hydrogen bonds should form¹. Nobody had ever captured a

picture of these electronic interactions before. Qiu's image seemed destined to become iconic.

"I was seduced by the image," says Mervyn Miles, a physicist at the University of Bristol, UK, who visited Qiu's laboratory at the Chinese National Center for Nanoscience and Technology in Beijing in September 2013, at the same time as the images were published online in *Science*. He recalls feeling both amazement at the finding and uncertainty that atomic force microscopy could really capture such a level of molecular detail. "You go, 'Oh, I'll think about that later,'" Miles says. More than four years on, researchers' awe has turned to scepticism. And there lies the challenge of using atomic force microscopy to capture images at subnanometre resolution.

The method's sensitivity to the electron density that surrounds atoms and exists

between them, including the bonds that hold them together as molecules, enables researchers to probe chemistry's fundamental secrets. "If I have just a single molecule of something, I can — in principle, at least — record the structure of that," says Peter Liljeroth, a physicist at Aalto University in Espoo, Finland. "There is no other experimental technique currently able to do this."

But this sensitivity can also produce artefacts in images, which means that scientists must take extreme care to show that what they see represents a real structure. And Liljeroth says that the subnanometre techniques he uses currently are "too complicated and too time consuming" to be adopted widely. He and other researchers are therefore hopeful that automation and machine-learning algorithms will help to simplify the technique. "This, I ►

► see like an engineering problem,” Liljeroth says. “This can be done.”

TECHNOLOGY TO WATCH

An atomic force microscope is like an atomic-scale record player (see ‘Anatomy of an atomic force microscope’). It requires a probe that comprises a cantilever (the record player’s arm) and a tip (the record player’s stylus). Both cantilever and tip are usually made from silicon, and the tip is coated with diamond or a metal. Tracking the probe as it moves over the surface of a sample (the record) maps the forces that the tip experiences. Because the tip and sample can both attract and repel each other, with the attraction potentially locking the tip in place, scientists typically make the probe vibrate. As the probe approaches the sample’s surface, the force that the tip experiences changes the frequency at which the probe vibrates, which researchers measure by shining a laser on the cantilever.

Although atomic force microscopy can now image structures at atomic resolution, researchers had to overcome several hurdles to make this feasible in practice. For example, experiments must be carried out in an ultrahigh-vacuum chamber, where working with lasers can be difficult. In the early 1990s, Franz Gießibl, while at Park Scientific Instruments (now Park Systems) in Sunnyvale, California, modified silicon cantilevers so that vibrational changes altered their electrical properties. This enabled him to convert changes in frequency of the probe’s vibration into electronic signals, making the use of lasers redundant.

Another problem was that silicon is relatively floppy. The vibrations of a silicon cantilever can span around 20 nanometres, a distance far greater than the diameters of the atoms that researchers want to observe. Needing a stiffer material, in 1994, Gießibl, who by then was living in Munich, Germany, turned to a surprising solution: the quartz tuning forks with which modern clocks keep time. Using a borrowed

atomic force microscope that he set up in his flat, Gießibl found that the tuning forks from Swatch wristwatches were “almost ideal in stiffness”, he recalls, because they vibrate on the subnanometre scale. Quartz, a piezoelectric material, can turn changes in frequency into electronic signals in a similar way to Gießibl’s modified cantilever. In 1996, Gießibl began to produce the qPlus sensor, a cantilever substitute that incorporates a quartz tuning fork. Now at the University of Regensburg in Germany, he licenses the associated patents to six companies that manufacture similar sensors with custom-made tuning forks.

One further advance was required to make imaging at atomic resolution routine. Because the shape and makeup of probe tips varies, it is difficult to precisely measure what they are doing. Tips can also move samples during scanning, leading to blurred images. To resolve these problems, scientists at IBM Research in Zürich, Switzerland, borrowed from previous work on scanning tunnelling microscopy. Similar to atomic force microscopy, the method uses

a probe to scan a surface, passing an electrical current through the sample to measure changes in conductivity rather than force. In the 1990s, researchers at Cornell University in New York affixed a single molecule of carbon monoxide to the scanning tunnelling microscope’s probe to create an atomically sharp tip. In 2009, the team at IBM applied the same approach to atomic force microscopy, producing striking images of pentacene, a chain of five interlinked six-membered carbon rings². The bonds between atoms appeared as clearly as if drawn in ink.

Carbon-monoxide tips are “very special”, says IBM’s Leo Gross, lead author of the 2009 paper. The molecule comprises just one carbon atom and one oxygen atom, with the carbon

flush against the tip and the oxygen positioned below like an atomic-scale stylus. As the tip approaches the sample, the atoms being studied repel the oxygen atom, causing the carbon monoxide to tilt to the side.

The resulting images show “very sharp bonds and you can see a lot of detail,” Gross says. If there is a ridge of electron density — like that associated with a chemical bond — tilting helps the atomic force microscope to obtain a clear contrast with the background. Indeed, that is precisely what Qiu saw for 8-hydroxyquinoline, with the structure’s carbon–carbon and carbon–nitrogen covalent bonds depicted in sharp relief. Yet tilting also creates “apparent bonds that you see sometimes where there are none”, Gross warns — for instance, between molecules. And in 2013, Liljeroth and co-workers suspected that Qiu’s team had imaged such a phenomenon.

Indeed, almost a year to the day on which Qiu’s paper was published online, Liljeroth’s team presented atomic force microscopy images that contained lines connecting atoms that do not interact through hydrogen bonding³. They suggested that such lines might be artefacts induced when neighbouring atoms nudge the carbon-monoxide tip, creating contrast where little or no electron density exists.

Other researchers interviewed by *Nature* say that hydrogen bonds are likely to make little contribution to what Qiu’s team observed. Gießibl, for example, thinks that the lines are “probably not hydrogen bonds”. However, “that’s not trivial” to conclude, he adds. “I had seen the paper when it came out and it sounded legitimate to me.”

Qiu says that what the images actually show is still up for debate. Both hydrogen bonding and nudges from nearby atoms might be involved, and when his team’s paper was published, the effect of tip tilting with carbon-monoxide tips hadn’t been studied fully. “Nobody, so far, found a real benchmark that is clear enough to distinguish the two,” he says.

“This is a learning field and we’re getting better at it.”

High-speed image collection

Recording conventional atomic force microscopy images can take more than a minute. But some microscopes can record images in less than a second, even at nanometre resolution.

To achieve such speeds, researchers such as Toshio Ando at Kanazawa University, Japan, use extremely small cantilevers (around 7 micrometres long), which resonate at very high frequencies. Standard cantilevers can exceed 100 micrometres in length. Using this approach, Ando’s team has been able to watch the motor protein myosin V ‘walk’ along a strand of actin at the rate of up to 20 images per second⁶.

At the University of Bristol, UK, Mervyn Miles’s team uses a quartz tuning fork, similar to that found in the cantilever substitute qPlus sensor, as a vibrating microscope stage for mounting samples on during scans⁷. The microscope’s probe initially rests on the sample, which is covered by water. As the stage vibrates more rapidly, the tip lifts off and ‘surfs’ over the water, explains team member Rob Harniman. This enables the researchers to scan samples at more than 100 images per second and up to around 1,300. “We’ve watched antimicrobial peptides making a hole in a membrane,” Miles says. **A.E.**

EXORCISING PHANTOMS

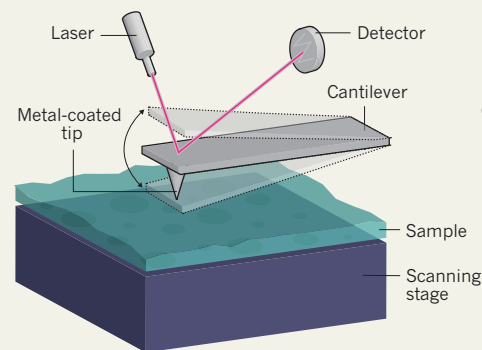
Such uncertainty is not a fatal flaw, Gross says. “This is a learning field and we’re getting better at it.” But how can researchers ensure that they aren’t misinterpreting what they see? One option is to anticipate results in advance of image collection. Pavel Jelinek from the Czech Academy of Sciences in Prague and co-workers have developed computational methods that reliably predict what scientists using carbon-monoxide tips should observe for many molecules⁴. Being able to make such a comparison is important because molecular images can be so compelling, explains Liljeroth, who used the simulations in his 2014 paper³. “If you see it, it must be real,” he says. “That’s the gut reaction.” But thanks, in part, to the computer simulations, scientists can now better anticipate the method’s idiosyncrasies.

ANATOMY OF AN ATOMIC FORCE MICROSCOPE

Atomic force microscopy (AFM) images the topography of a material by dragging an atomically sharp vibrating probe across its surface. Advances in probe design are sharpening the method's resolution.

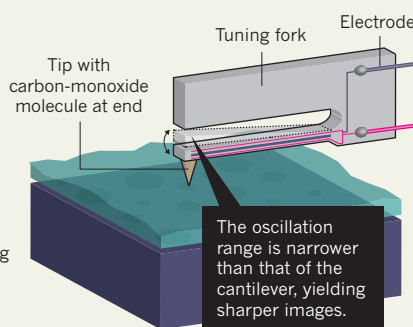
CONVENTIONAL AFM

The probe tip is affixed to a flexible silicon cantilever, the deflection of which is tracked using a laser.



SUBNANOMETRE-RESOLUTION AFM

A vibrating quartz tuning fork is used in place of a cantilever. Changes in its oscillation frequency are detected as electrical signals.



microscopy images", Liljeroth says. Although some level of expertise will always be necessary, he hopes that machine-learning algorithms and automation will help to reduce the time that is needed for tip preparation.

Moriarty echoes this sentiment, as preparing a good tip can require a lot of trial and error. The only way of knowing whether a researcher has attached the correct molecule or atom is to record a clear image. The researcher must then repeat the image-collection process with fresh tips, to gather enough observations to be confident in his or her interpretation. Moriarty admits that scientists can find this a "soul-destroying" practice. "Automating that process is the way to go," he says.

Moriarty and others have already taken a step in this direction by developing an automated process for using tips to move hydrogen atoms⁵.

Despite slow progress, Moriarty finds cause for optimism in astronomy, a field in which machine learning is helping researchers to judge when data might be meaningful. The Laser Interferometer Gravitational-Wave Observatory (LIGO) collaboration, for instance, successfully applied computational algorithms to distinguish possible gravitational waves from artefacts.

But limitations remain, Moriarty admits, because even LIGO needs scientists to confirm manually whether potential detections are real. Tip preparation at the push of a button would require a system to be able to accurately judge the quality of the images that it gathers, he says. "The only way to automate is to train the blasted machine to recognize when it's got a good image," Moriarty says. "If a machine could do that, those astronomers would be out of a job." ■

Andy Extance is a freelance writer based in Exeter, UK.

1. Zhang, J. *et al. Science* **342**, 611–614 (2013).
2. Gross, L., Mohn, F., Moll, N., Liljeroth, P. & Meyer, G. *Science* **325**, 1110–1114 (2009).
3. Hämäläinen, S. K. *et al. Phys. Rev. Lett.* **113**, 186102 (2014).
4. Hapala, P. *et al. Phys. Rev. B* **90**, 085421 (2014).
5. Möller, M. *et al. Nanotechnology* **28**, 075302 (2017).
6. Kodera, N., Yamamoto, D., Ishikawa, R. & Ando, T. *Nature* **468**, 72–76 (2010).
7. Picco, L. M. *et al. Nanotechnology* **18**, 044030 (2007).

CORRECTIONS

The Toolbox article 'The future of scientific figures' (*Nature* **554**, 133–135; 2018) implied that Benjamin Delory developed the persistence barcode method. In fact, he and his colleagues developed an analysis pipeline that relied on and adapted an existing method.

The Technology Feature 'Deep learning for biology' (*Nature* **554**, 555–557; 2018) erroneously affiliated Mark DePristo at Verily Life Sciences. He is, in fact, at Google. Also, the DeepVariant tool was developed jointly by Verily and Google.

Liljeroth's team has used carbon-monoxide tips to study electronic components embedded in graphene nanoribbons. Yet current approaches cannot simulate non-flat structures well enough to help in their interpretation. Together with his Aalto colleague Adam Foster, Liljeroth is developing machine-learning algorithms and artificial-intelligence programs to predict images of objects of any size, configuration or orientation. The pair have assembled a network of collaborators, and are seeking funding to support the effort.

Ultimately, this approach could lead to fully automated data interpretation, Liljeroth says. But Philip Moriarty, a physicist at the University of Nottingham, UK, suggests that this is unlikely to work in all cases, because even experienced researchers disagree on what atomic force microscopy data show. Moriarty cites the results of a recognition test in which his team was asked to classify images into categories — such as whether they showed atoms assembling individually, in rows or as pairs. The highest scoring participant succeeded only about 70% of the time. "If humans can't recognize one image from another, we've got a bit of a problem," Moriarty warns, because researchers' judgement provides the benchmark by which algorithms are trained. Yet by participating in Liljeroth's network, Moriarty hopes to explore the possibility that, with access to appropriate image data, artificial-intelligence-based classification systems could outperform people.

But there's little image data available on which to train such systems, Gross says. Only about 100 known molecules have been resolved to atomic resolution using carbon-monoxide-tip atomic force microscopy, he estimates. Although automated classification should be tried "at some point", Gross thinks that it's too early, at present.

Simulations such as Jelinek's could provide a suitable training set, Liljeroth suggests. "The

question is whether these synthetic images are close enough to experiments."

SOUL-DESTROYING REPETITION

If nothing else, automation could help to mitigate the arduous practical challenges faced by researchers who use atomic force microscopy (see 'High-speed image collection'). Filipe Junqueira, a PhD student in Moriarty's lab, is studying how to produce arrays of thin columns of gallium arsenide. Each column, known as a nanowire, has a diameter of 10 nanometres or greater and is grown inside a stainless steel ultrahigh-vacuum chamber. To image the nanowires, Junqueira must overcome practical obstacles such as experimental noise and sample manipulation using a metal arm known as a wobble stick. His measurements contain interference that might be related to construction work being carried out several hundred metres away, even though the atomic force microscope he uses is housed in a basement and is supported by a table that can dampen vibration.

Other labs have taken more drastic steps to minimize the effects of noise. At Vienna University of Technology, Ulrike Diebold's team suspends its microscope from 36 vibration-damping elastic cords. When combined with an automated system that keeps the system level, this enables carbon-monoxide-tip atomic force microscopy that provides "beautiful images", according to team member Martin Setvin.

And there are further challenges. To mount a carbon-monoxide molecule on a tip, researchers must push the microscope's probe up to a surface coated with carbon monoxide, and then pass an electric current through the probe. It can take hours to get the process right, Setvin says. "If you lose the carbon-monoxide molecule, you have lost a day of work."

For a PhD student with little experience of scanning-probe techniques, it can take "a couple of months to start getting very nice atomic force

CAREERS

CHILLY CULTURE Male environment linked to dropouts by female researchers **p.551**

CONFERENCE COMFORT Female scientists call for child-friendly settings **p.551**

BLOG Personal stories and careers counsel <http://blogs.nature.com/naturejobs>

STEPHAN ELLERGMANN/LAIF/EYEVINE



Scientists at the biotechnology company Affymetrix in Santa Clara, California. Positions in the biotechnology sector are highly sought after.

TRANSITIONS

Industrial strength

Life-science industry insiders offer tips on how to beat job competitors.

BY KENDALL POWELL

When immuno-oncologist Martijn Bijker decided to move from academia to industry, he asked a friend to review his CV. His friend — who had worked in the pharmaceutical sector for two decades — told him to relegate his lists of publications, posters and presentations to the back of the document. The focus, he said, should be on teamwork skills and the ability to perform the job requirements.

The advice was an eye-opener for Bijker. He was used to the academic hierarchy, which values individual achievements above all else. But in corporate research, candidates must focus on collaboration.

Bijker subsequently founded From Science to Pharma, a consultancy and recruitment firm

based in Sydney, Australia. He now regularly issues similar counsel to scientists hoping to shift from academia to industry, encouraging them to flip their CV — and their mindset — to reflect the differences between the sectors.

Biotechnology and pharmaceutical industry positions are attractive to many recent PhD graduates — and are thus highly competitive. More than half of the respondents to *Nature's* 2017 Graduate Student Survey said that they would like to work in industry, and nearly one-quarter said an industrial position was what they most wanted.

Researchers who seek these jobs must switch from touting their publications, awards and citations to highlighting the value they bring to a team effort. They also need to identify their vocational strengths and talents, such as communication and management

skills that transcend bench-top prowess and technical proficiency.

THE SWITCH TO INDUSTRY

Junior researchers who are considering industry should aim as soon as possible to develop the skills and expertise that will be attractive to hiring managers, says Alaa Abdine, a PhD-trained biophysicist who is now an executive recruiter for Crossover Search in New York City, which specializes in biotechnology companies. In last year's graduate-student survey, just 20% of respondents had sought the advice of a researcher who held the job they wanted, and only one-third received input on non-academic careers from their advisers. "Don't wait until your eighth year of a postdoc and then come to a recruiter looking for a job," says Abdine. ►

► Those who have made the switch to industry recommend that researchers identify their goals and interests before they launch a job search. After applying for a few different types of job, Forum Raval — who has a PhD in immunology and virology — realized that she really wanted to do translational research, developing products or procedures that would help people.

When discussing a friend's job in genetic testing, she realized that it fitted with her vision of translational research, even if it did not perfectly match her academic background. She managed to get an interview with Quest Diagnostics, a health-care diagnostic firm in Marlborough, Massachusetts, despite lacking solid genetics credentials on her CV. In the interview, she banked on selling the skills she had obtained during her PhD, which required her to adapt and switch between complex tasks. "I was sweating the whole time," she recalls.

She got the job. Now, as a scientist at Quest, she analyses data from people with hard-to-diagnose disorders, to determine whether the disorders have a genetic basis and, perhaps, a therapeutic solution. "I'm loving it," she says. "I'm still using my brain for science and directly helping people."

Researchers mulling over the switch to industry should also consider whether they are driven by the singular pursuit of their own ideas. "This is more difficult to do in industry, where research goals are usually set more narrowly," says Praveen Kumar Vemula, a biomaterials scientist at the Institute for Stem Cell Biology and Regenerative Medicine in Bangalore, India. Although he has remained in academic research, where he has the freedom to develop his own ideas, he has deep ties to India's biotechnology industry and has spun off several companies and licensed his technologies to others. To excel in industry, he says, means collaborating with others.

DO YOUR HOMEWORK

Job seekers must show a deep understanding of the company that interests them, such as its products, workplace culture, history, competitors, business-development plans, regulatory challenges and how it fits into the marketplace. They must also know a great deal about the position they want. Martin Rees, who is now a regional medical director in the Sydney office of biopharmaceutical company Stallergenes Greer, gave himself six months to learn everything he could about becoming a medical science liaison (MSL) a couple of years ago. An MSL is a scientific expert in a pharmaceutical or biotech company who liaises between the firm and the medical and scientific professionals who use its products.

He met informally over coffee with MSLs from different businesses and with marketing and salespeople from those companies to learn about the employers' cultures, products and competition (see P. Fiske *Nature* **538**, 417–418; 2016). "You have to do your homework. That's

APPLICATIONS AND INTERVIEWS

Get yourself prepared for a transition to industry

Shine on paper

- Review the CV of someone with the job you want or who works at the company you're targeting. Note the formatting and length.
- Put long lists of publications, posters and presentations at the end of your CV or don't include them at all.
- Consider including a section such as 'Most Relevant 5 out of 20 Publications', so that employers are able to see your overall productivity without having to wade through a long list.
- Use active, past-tense verbs such as 'implemented', 'designed' or 'developed' to convey results accomplished.

Be polished in person

- During an interview, you should back up statements about your skills, whether hard or soft, with real-world examples from your PhD or postdoctoral work. Don't discuss other job options or interviews with anyone on the interview team. Here are a few examples of 'curve ball' interview scenarios.
- Give an example of a time you overcame a



Industry-ready CVs need a targeted approach.

frustration with a colleague.

- Explain how you handle criticism.
- Tell us, why should we hire you rather than someone else?
- Talk about a time you failed.
- Take a timed test to write a short data-analysis computer program. **K.P.**

really what's going to get you through the door for an interview," says Rees.

Professional associations can help to provide an inside view of companies' operations. The chapter of the Association for Women in Science in San Diego, California, for example, offers tours of local biotechnology and pharmaceutical companies for its members.

Job seekers should also spell out to hiring managers and other interviewers exactly how they will fit into a prospective employer's portfolio and mission. In 2017, the scientific recruitment company C-Drive in Bangalore conducted a survey of Indian companies in the life and chemical sciences, and it found that firms looking to hire PhDs and postdocs are especially frustrated by the candidates' lack of awareness about the company and its scientific context. If a cell biologist, for example, is vying for a job at a biotech firm or drugmaker that specializes in cancer treatments, the candidate should show in their application how their background fits with the target company's drug or molecular pipeline. "Many times there is a connection," says Shyam Suryanarayanan, who is chief executive of C-Drive and founded a partner company that provides industry-readiness training. "But it's not up to the employer to try to figure it out."

INDUSTRY CONTACTS

In the hyper-competitive environment of industry, getting a foot in the door can often depend on a researcher's contacts in a sector or particular company. Barbara Preston,

co-founder of recruiting firm PharmaScouts in San Diego, recommends that researchers build their networks by using LinkedIn to contact scientists who share research interests, backgrounds or even hobbies. If the junior researcher is attending an upcoming conference that the company scientist is also likely to attend, it's ideal to get in touch beforehand and ask to meet over coffee. "Use your position as a graduate student or postdoc as a plus, not a minus," Preston says, adding that it's important to be honest and direct but not to ask for a job. "Say, 'I'm coming up to a decision point in my career, and I want to make an informed decision. Can I ask you a few questions about your experience in industry?'"

Even researchers who have high-impact publications should seek contacts who work for their target companies. Those contacts could offer personal introductions, or they could put the researcher's name forward to set them apart from competing applicants. "Networking is the number-one thing anyone needs for getting any job," says Abdine.

Graduate students and postdocs should aim to find an academic adviser who has links to local companies, as Vemula does. It can also be helpful to develop a connection with lab heads in the student's or postdoc's speciality who have spun off their own business, because they are also likely to have industry contacts. Vemula, for example, encourages his industry-interested lab members to help prepare for and sit in on the biweekly or monthly meetings

with his industry collaborators to discuss a project's progress and timelines. "Executing projects is very different with industry partners, and the lab members get first-hand experience in what an industry job will be like," says Vemula.

Other, more formal initiatives help junior researchers to connect with industry leaders. Australia launched a programme in 2017 to encourage doctoral students in the sciences to team up with industry-based mentors over the course of a year. Fourteen of the nation's 40 universities participate in the Industry Mentoring Network in STEM (IMNIS) programme, which matches about 200 science students with mentors. The pairs meet monthly, and students may get a workplace tour or invitation to industry networking events, for example. "It's a great opportunity for open conversation at an early stage of the student's career and gives students a chance to expand their networks," says Marguerite Evans-Galea, IMNIS executive director.

Evans-Galea says that, in Australia, nearly 70% of industry positions are not advertised because they are filled through word of mouth. Abdine says the same about entry-level research positions in the large US biotechnology hubs of San Francisco and San Diego in California, and Boston in Massachusetts.

CAREER KEYWORDS

Researchers who want to transition into industry will need to convert their academic CV into one that is specifically appropriate for industry. They should also prepare for an interview process that is different from that in academia (see 'Get yourself prepared for a transition to industry').

Industry insiders suggest that the researcher write an eight- to ten-bullet-point summary at the top of their CV that highlights their training, background, career goals and skills, using keywords that will grab the attention of the first screener — probably a human-resources officer who may not have a scientific background. Those keywords should be the same ones that the hiring company used in its job advertisement, particularly in the sections that discuss the job's requirements and the company's research goals. "This is like the abstract of a paper — is this interesting enough to read the rest of the paper? If not, your CV goes into the trash pile," Bijker says.

An industry CV, known in the United States as a résumé, should reflect the candidate's productivity, communication skills and innovation, says Preston. "Did you initiate a project, implement a new technique, or optimize an assay? What was innovative about your research? What did you shed light on to change scientific understanding?" She recommends the site www.scientificresumes.com, a service developed by scientists to help

researchers convert academic CVs into those appropriate for industry.

The interview itself is likely to include queries on workplace behaviour. And candidates should prepare for unexpected requests, such as "Tell me a joke". Preston says that these 'curve ball' scenarios can help screeners to identify candidates who can think quickly or who might have inappropriate biases.

Interviewees might also receive a research case study or roadblock to solve on the spot. Raval says that, during an interview, candidates should be comfortable admitting if

"You have to do your homework. That's really what's going to get you through the door for an interview."

they don't have an answer, but should also use the opportunity to explain their thought process. "Saying 'I don't know, but my guess would be ...' is a very hard thing to say in an interview," she concedes. "But if you show them how you think, sometimes that's more informative than the actual answer."

Most final shortlisted applicants will be asked to give a scientific presentation. Preston recommends that candidates ask the hiring manager what types of people will be in the audience and what they would like to hear. "Be prepared to tell a story," she says, adding that candidates should remember that important audience members are not likely to be technical specialists, so talking points and slides need to be accessible. Employers will be listening for a candidate's research thought process and for their communication and working style, Preston adds. "The presentation can make or break you," she says.

Recruiters and hiring managers stress that junior researchers need to prepare immediately once they decide to transition to industry. "A lot of people have done a PhD and gone on to a postdoc by default," says Suryanarayanan. But, he says, those researchers are often only delaying the inevitable. "You should figure out who you are," he says. "It's good to think about what you want to do and what you are good at — but think about it before you feel cornered into an existential crisis." ■

Kendall Powell is a freelance science writer based in Lafayette, Colorado.

CLARIFICATION

The Careers feature 'The write stuff' (*Nature* **555**, 129–130; 2018) should have made clear that Altmetric is part of Digital Science, a company owned by Holtzbrinck Publishing Group, which is also the majority shareholder in *Nature's* publisher, Springer Nature. Nature Research Editing Services is also owned by Springer Nature.

CHILLY CULTURE

Women forced out

A male-dominated workplace and a competitive culture that often shuts out family life might be contributing to the decline in the proportion of women at successive stages in research. In a qualitative study of 28 US female PhD students in physical sciences and engineering, led by Bianca Bernstein at Arizona State University in Tempe (M. Cabay *et al. Soc. Sci.* **7**, 23; 2018), 12 said that they did not want to pursue research careers. Of those, 6 blamed their workplace environment and culture, including 2 who said they could no longer work within a male-heavy profession. Some of the female students reported that they felt ignored, dismissed or excluded from scientific conversations and other lab interactions among male colleagues. One said that a male colleague attributed her winning a scholarship to her gender and to quota filling. Some of the female students also reported being asked disproportionately often to perform 'women's work', such as cleaning up the lab or performing clerical duties.

DIVERSITY

Open doors to children

Conference organizers must make their events more welcoming and accessible to parents of young children, say 46 scientist-parents. Rebecca Calisi, a behavioural neuroscientist at the University of California, Davis, and her colleagues offer a blueprint for improving attendees' experience (R. M. Calisi *et al. Proc. Natl Acad. Sci. USA* <http://doi.org/ck8h>; 2018). By not accommodating children, the authors say, conferences can unintentionally create barriers that exclude large numbers of scientists — especially mothers at an early stage of their career who might not be able to afford childcare. "One part of promoting diversity is supporting women with children," Calisi says. Rules about children seem to change from conference to conference and even from hour to hour, says Calisi, who notes that researchers with babies were barred from a poster session at a large conference last November, even though the official policy permitted children in the exhibition area. A practical, comfortable space for breastfeeding or pumping breast milk is an important provision, Calisi says. The Society for Neuroscience, for one, aims to become more inclusive. "The society is exploring ways to enhance the spaces for nursing mothers," says spokesperson Kara Flynn.

THIS BIG

A ray of hope.

BY JOHN COOPER HAMILTON

I should kill them all... What? We're transmitting?

Greetings, fellow researchers, wherever you are. Important news! In my laboratory earlier this week. I — *Really? Oh, very well.*

I have just been advised to begin with a joke. *This card?*

What is red — this is a riddle — what is red shading gradually to a frothy transparency and travels at 4,300 r.p.m.? Answer: a graduate student in a Pelsen 557VES centrifuge! But you can only fit them in 300 millilitres at a time! Pause for laughter.

Laugh, damn you, laugh! Fire will rain down on your cities and your sewers will fill with a million angry — *Now what? Oh. Of course, you're correct.*

Ahem. Yes. What happened in my laboratory was a third case of spontaneous human combustion, henceforth SHC.

Funny story: it was Osgood, one of my graduate students. Osgood, Max and I sat down to eat our bag lunches in the break room. I had onomi sashimi, which is fresh blue-whale meat. I know, I know! It's terrible, but I have it with a side of brown rice rather than white. That's healthier. Max had a banana, as is traditional. Osgood pulled a thermos of coffee from his bag, said, "Would you like some?" and burst into flames.

Oh, how we laughed!

Well, I did. Max scampered off for a fire extinguisher and Osgood rather uselessly thrashed about and screamed. He lost control so far as to approach me, at which point I shot him with the harpoon gun handily sitting nearby. I directed Max to extinguish the flames, but only after preparing an impromptu banana flambé.

A recipe shall follow this broadcast.

Max said, "That was the second this week." I said, "No, dear friend, that was actually number three." *Ouch.*

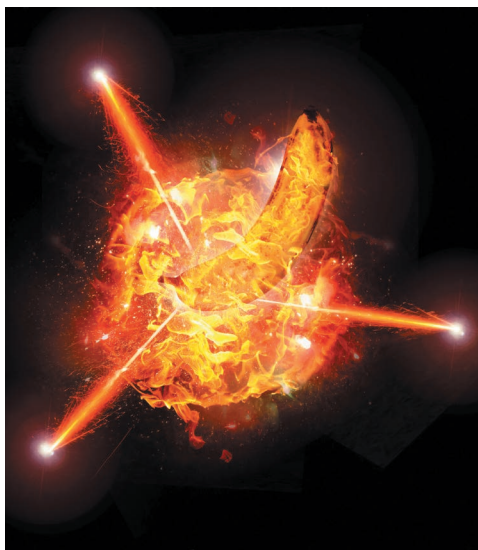
Excuse me. *I* said, "Max, that was the second this week." He said, "No, the third."

I enjoy mentoring, but with Osgood's exit my island was out of mentees. Before obtaining more, it behooved me to ascertain the cause of Osgood's death. The cause of SHC, I mean. The cause of Osgood's death was a harpoon in the head.

My first thought

was a matter/contra-matter reaction — I had several difficulties repurposing my

largest collider into a cappuccino machine. But a little noodling demonstrated that a reaction would have resulted in an explosion of approximately 3,140 megatons. Or, actually, 3,129, deducting the 300 ml of matter used in the previous day's experiment.



Such an explosion would have utterly destroyed my secret island base. It might even have cracked a tectonic plate, but that's a thought for another day. The point is, it wasn't contra-matter.

Other SHC investigations:

My new hot sauce. False: 50% toxic dose at 2.1 grams, but no amusing side effects.

The British super-spy in the holding cell. False: he'd escaped and was no longer on the island.

The experimental RF resonant cavity thruster, or 'EMDrive'. False: it's identical to the model being debunked by NASA, and it does absolutely nothing. I also checked my in-use EMDrive and observed nothing unexpected.

I admit, fellow researchers, that I despaired and took to the consolation of work. This month's project was developing a new neutrino detector. Traditional models, created by traditional so-called scientists at their hide-bound universities where they will *rue* the day! Rue the day they threw me out and called me mad! Me? Mad!!! *Ouch.*

Thank you, Max.

As I was saying, current detector designs use mineral-oil-filled tanks, hundreds of metres long, to detect the elusive particles. Ha! Again, I say ha!

It was — and is! — my theory that, owing to the principle of *resonance*, a tank the same size as a neutrino will detect particles just as efficiently. Dozens of mineral-oil-filled tanks were arranged before my shrink ray, a standard RH 1050 model, more than capable of reducing a tank to subatomic size. A laboratory fire — graduate student number 1 — had disrupted the procedure and the tanks now varied in size. The one before me was something over two metres in diameter at the moment it exploded.

I was lying right here, on the very infirm bed from which I now broadcast, when I awoke and said, "Max, I've got it! The resonance theory is right, and contact is energetic enough to translate the neutrino's mass into heat!" *Ouch. Damn your furry little hide, Max.*

I was lying right here on the infirm bed when I awoke, and Max said, "Doc, I've got it! Your resonance theory is right!" Et cetera.

While I was unconscious, Max, a very capable assistant, I freely admit — *Now put that damned knife down* — achieved confirmation of the theory. Neutrinos are really *this big*.

I'm holding my arms all the way apart. That big.

Need I remind my august peers that extraordinary claims require extraordinary evidence? The idea that the entire range of subatomic particles is *actually* vanishingly small doesn't pass the mad-laugh test. It was an assumption made by our small-minded brethren. They forced us into hiding, into — outside the annual conference at Cabo — communicating only via scrambled radio transmissions.

They shall pay dearly for their lack of vision.

Neutrinos are not small and common. They are large and rare. Detector tanks shrunk to exactly one and a half diameters refracted neutrinos through my laboratory at a furious rate, causing SHC. Focused, the neutrinos create lethal beams capable of spanning any distance, penetrating any armour.

Rise, brothers and sisters, rise! I've dibs on the basic deathbeam patent, but there shall be enough death, and enough beams, for all! ■

John Cooper Hamilton writes humorous genre fiction, except when it's sombre or creepy. He even writes literary fiction if he thinks he can get away with it.

ILLUSTRATION BY JACEY

nature INDEX 2018 JAPAN

NATURE, VOL. 555, ISSUE NO. 7697 (22 MARCH 2018)

STALLED AMBITION

A sense of crisis quickened by China's stunning rise has surrounded Japanese science since Nature Index's 2017 Japan supplement highlighted the country's faltering research performance.

This year we report continued decline: a further 3.7% drop in Japan's contribution to high-quality scientific output in 2017 deepens the 19.6% decrease recorded between 2012 and 2016. It doesn't bode well for "Society 5.0", the nickname for Japan's ambition to become a 'super-smart' society. But without the budgetary or human resources to match China, how to stop the drop?

One answer is to boost research productivity. Japan has plenty of scope for that, as the normalized metric introduced in this supplement reveals. Looking at performance relative to the volume of publication output gives a clearer picture of the comparative

effectiveness of Japan's research effort. Normalized weighted fractional count is derived by dividing a country or institution's WFC in the Nature Index by its total number of natural sciences articles in Elsevier's Scopus database. On this measure, Japan overtakes the United Kingdom to move from 5th to 4th place among the world's leading producers of high-quality scientific output in the index.

The bad news for Japan is that it performs worse than 29 other developed economies in the amount of high-quality research it produces per R&D dollar invested. It remains to be seen whether the reforms described in this issue will yield strong enough results for Japan to remain a leading science nation, let alone become the 'super-smart' society of its aspirations.

Catherine Armitage
Chief editor, Nature Index

Resistance to reform

S52: Moves to make Japanese universities more agile and dynamic are being met with a stubborn adherence to tradition.

Calling Rikejo

S59: The under-representation of women in Japanese science will not be solved without a change of general work culture, says neuroscientist, **Noriko Osumi**.

Partners in discovery

S60: Osaka University is pursuing links with industry to fund the research meant to shore up its future.

Strength from weakness

S61: Plans for a next-level society will demand a much bigger pool of expertise to bring the real and cyber worlds together.

Pillars of a smart society

S62: Meet some Japanese scientists whose work in bionics, autonomous cars, and cyber-commerce, exemplifies the nation's planned trajectory.



Facing down disaster

S66: The frequency of natural disasters makes research into warning and mitigation an obvious national priority.

Short-term generation

S67: Almost half of early-career researchers can only hope for a temporary employment contract, so the government wants companies to step in to improve their prospects.

The tables

S68: How Japan's institutions and companies stack up in the Nature Index.



ON THE COVER

Kanako Harada's bionic humanoid is among the inventions aimed at transforming Japanese society.

EDITORIAL: Catherine Armitage, Smriti Mallapaty, Stephen Pincock, Rebecca Dargie, Herb Brody, Victoria Kitchener
ANALYSIS: Aaron Ballagh, Bo Wu, Willem Sijp **ART & DESIGN:** Alisdair Macdonald, Kate Duncan, Mohamed Ashour, Ruffi Lu, Chika Takeda, Wojtek Urbanek **WEB DEVELOPMENT & DESIGN:** Bob Edenbach, Olivier Lechevalier, Roberto Espinoza, Naomi Nakahara, Erika Suzuki **DATA QUALITY:** Jörn Ishikawa, Yuxin Wang, Paul Glaeser, Randell Roach, Miho Kawana, Megha Katyal, Alexander Scherrmann **PRODUCTION:** Sue Gray, Karl Smart, Ian Pope, Dipti Shah **MARKETING:** Stacy Best Ruel, Angelica Sarne **PROJECT MANAGEMENT:** Richard Hughes, Shoko Hasegawa, Takeaki Ishihama, Kazuki Kurebayashi **SALES:** Maki Ishikawa, Tommy Yim **ART DIRECTOR:** Kelly Buckheit Krause **PUBLISHING:** Nick Campbell, Richard Hughes, David Swinbanks.

NATURE INDEX 2018 JAPAN

Nature Index 2018 Japan, a supplement to *Nature*, is produced by Nature Research, the flagship science portfolio of Springer Nature. This publication is based on data from the Nature Index, a Nature Research website maintained and made freely available at natureindex.com.

NATURE EDITORIAL OFFICES

The Campus, 4 Crinan Street,
London N1 9XW, UK
Tel: +44 (0)20 7833 4000
Fax: +44 (0)20 7843 4596/7

CUSTOMER SERVICES

To advertise with the Nature Index, please visit natureindex.com/client-services-feedback@nature.com. Copyright © 2018 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.

IRWIN WONG

KENJI SUZUKI

Tokyo Tech's president, Yoshinao Mishima, is determined to lure world-class students to the institution, and is prepared to make sweeping changes to make it more attractive.



RESISTANCE TO REFORM

Efforts to improve Japan's scientific status include top-down institutional change, but tradition is hard to break.

BY ICHIKO FUYUNO

In April 2016, Tokyo Tech's president, Yoshinao Mishima, made the unprecedented move of appointing department heads himself. For 137 years, positions at the Tokyo Institute of Technology had been decided by a faculty vote and a routine rotation based on seniority. Mishima's actions rattled the administration, but they weren't arbitrary. Revised laws governing national universities gave presidents greater discretion. The aim is to shake up the slow, consensus-based decision-making of Japanese university management, and to allow swift reform to the way universities are run.

Changing the culture of universities is central to Japan's efforts to arrest the deterioration in its research performance. Last year alarm bells rang over a 19.6% fall in its high-quality scientific output, the sharpest fall among the 10 leading research nations, as measured by the Nature Index over five years to 2016. There was a further 3.7% decline in 2017.

It is difficult to recalibrate the research ecosystem, and it will take time, as the situation at Tokyo Tech illustrates. Mishima is an exception: many leaders at large universities are slow to take advantage of the governance reforms.

In addition to appointing directors, Mishima also removed the final say on recruiting new

researchers from faculty councils, and put in place a human resources committee with the president at its helm. "We have to make the university more attractive to lure the world's top-class students," says Mishima, who wants Tokyo Tech to join the ranks of the best engineering colleges globally by 2030.

It's a tall order. The institution currently ranks at 65th for engineering and technology in the Times Higher Education's World University Rankings 2018. In the Nature Index, it ranked 79th globally for its total contribution to paper authorship in 2016.

But on other measures that matter for Mishima, Tokyo Tech stands out. The university ranks 9th among Japanese academic institutions above big names such as Keio, Tohoku and Kyushu, when its contribution to the index over the past six years is considered as a proportion of its total natural sciences output in the Scopus database.

STAGNATED PERFORMANCE

Japan holds on to 5th place for output in the Nature Index after the United States, China, Germany and the United Kingdom. However, on one measure of efficiency — high-quality science article output per R&D dollar, measured by the share of authorship in the index against gross spending on research and development — it falls

to 30th place among the 42 economies for which OECD R&D spending data are available.

Universities are under pressure from the Ministry of Education, Sports, Culture, Science and Technology (MEXT) to improve management efficiency.

"Most universities fund researchers according to distribution rules with little flexibility, and do not evaluate their management strategies," says Atsushi Sunami, a science-policy expert at the National Graduate Institute for Policy Studies. He contends the biggest problems in national universities are inflexibility and the lack of agility in research and human resources management.

The management reform initiatives have yet to demonstrate any effect on national research outcomes. Some reforms have been met with stubborn resistance and adherence to tradition.

"They don't explicitly oppose me, but they are slow to take action," says Seiichi Matsuo, president of Nagoya University, describing faculty members' responses to his efforts to strengthen leadership. Nagoya University ranks 11th among Japanese academic institutions when normalizing its contribution to high-quality research against its total output in the natural sciences.

Yuko Ito, a policy expert at the Japan Science and Technology Agency (JST), says the extent of governance reforms varies by university. Until about five years ago, she heard professors

TOKYO TECH

TOKYO TECH

calling for a return to the old system. But she says a shared sense of crisis is changing attitudes.

National university reforms began in 2004, when MEXT gave university management greater autonomy to build their competitiveness. Public research institutes underwent a similar change beginning in 2001.

In 2010, the state redoubled efforts to overhaul the university governance system and make them more competitive globally. MEXT directed national universities to accelerate organizational reforms such as increasing the use of English, introducing flexible human resources and payroll systems, and building strong presidential leadership.

Mishima, for example, integrated undergraduate and graduate schools into six schools and 19 departments. He set up an international advisory board, and formed councils to smoothly implement reforms.

Mishima plans to raise the proportion of graduate school lectures given in English from 31% in 2014 to almost 100% by 2021. This change, he says, “will not only help attract more excellent overseas students, but allow Japanese students to get used to communicating in English before building a global career.”

But many disagree with the approach, arguing that poor English among some teachers and students could affect the quality of education. “Why can’t we use Japanese,” asks a researcher at Tokyo Tech who requested anonymity. He points out that in many countries, such as Germany and France, courses are taught in the native language.

Evaluation has also been overhauled. In 2016, MEXT made a tenth of the total management expenses grants competitive. National universities had to classify themselves into one of three groups: serving the local economy, contributing to global research, or a unique teaching focus. Dozens of numerical key performance indicators are used to assess their strengths in each group; funding is commensurate.

Some university leaders question the value of meeting the targets given the relatively small amounts on offer. Tokyo Tech, for example, received the highest possible score on its evaluation, yet the increase in its management

expense grants for 2017 was a mere 0.1%. Some believe the intensification of evaluations with their heavy burden of paperwork is becoming counter-productive. Ito says the new system is diverting energy from research. “The government’s evaluation is getting more severe in many ways. Universities look worn out from being evaluated.”

MORE WITH LESS?

Changes have coincided with a period of austerity. Following rapid increases in the 1990s, the pace of growth in R&D spending has stalled since 2001, due to social welfare costs during 20 years of economic stagnation.

Some blame a lack of money to employ adequate staff, more than management inefficiencies, for the decline in research quality.

Many universities have stopped hiring staff, replaced tenure positions with contract-based positions and cut basic research spending. At the top research universities, nearly 40% of researchers were employed on contracts in 2013, up from 27% in 2007, according to NISTEP.

Funding cuts have affected smaller, regional universities more severely, as they struggle to earn external funding compared with prestigious universities such as The University of Tokyo.

Kouichi Okunishi is a physicist investigating condensed-matter theory at Niigata University, and part of a group established in 2015 to share concerns about the university’s decision-making system. He is among those who say the lack of people is more debilitating than the lack of money for research.

In early 2016, when the university’s management suspended hiring for two years, the group collected 300 signatures on a petition. But the hiring freeze was put in place anyway, he says. As a result, about 20 tenure positions have been lost from the science, agriculture and engineering faculties due to staff members retiring, Okunishi says. Meanwhile, student numbers remain the same and faculty need to spend more time teaching undergraduate students and less on research.

Toshio Suda, director of the International Research Center for Medical Sciences



Researchers at Tokyo Tech.

(IRCMS), launched in 2015 at Kumamoto University, says recruitment is his biggest challenge. Although some 40% of IRCMS researchers are from other countries, Suda says it is very difficult to hire principal investigators from abroad, partly because there is little money for relocation. Instead, Suda is seeking to attract talented overseas postdocs and runs research intern programmes for students within Asia.

TRYING TIMES

The government has set a target of raising R&D investment from 0.65% to 1% of GDP by 2020. That translates to an additional 900 billion yen (US\$8 billion) over the next three years. The 2018 science and technology budget delivered a sharp increase of 250.4 billion yen in new money. A change in accounting method put a further 80.5 billion yen on the bottom line, bringing the total increase to 7%.

Analysts welcomed the new money, saying it put the government on track to meet the 2020 target. Three-quarters of it came from ministries agreeing to invest in new technologies and automation to improve the productivity of infrastructure and services.

But no increase was approved for universities, despite a request for a 4% rise of 1.14 trillion yen, says Toshinori Sano, deputy director of the National University Corporation Support Division at MEXT. Meanwhile the number of students on financial aid that universities have to support using their management expenses increased by a few thousand.

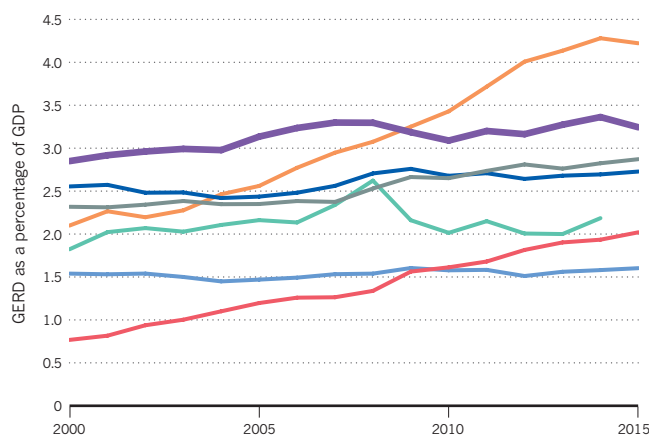
“Japanese science is still competitive,” says Guojun Sheng, an embryologist who joined IRCMS in Kumamoto in 2015. “But, now is the time to improve the efficiency and make the environment more attractive.” ■

SOURCE: OECD

SCIENCE SPENDING

Japan spent 3.29% of its GDP on research and development in 2015. This was a lower proportion than South Korea, but higher than the United States and China.

Japan
South Korea
United States
Germany
Singapore
United Kingdom
China



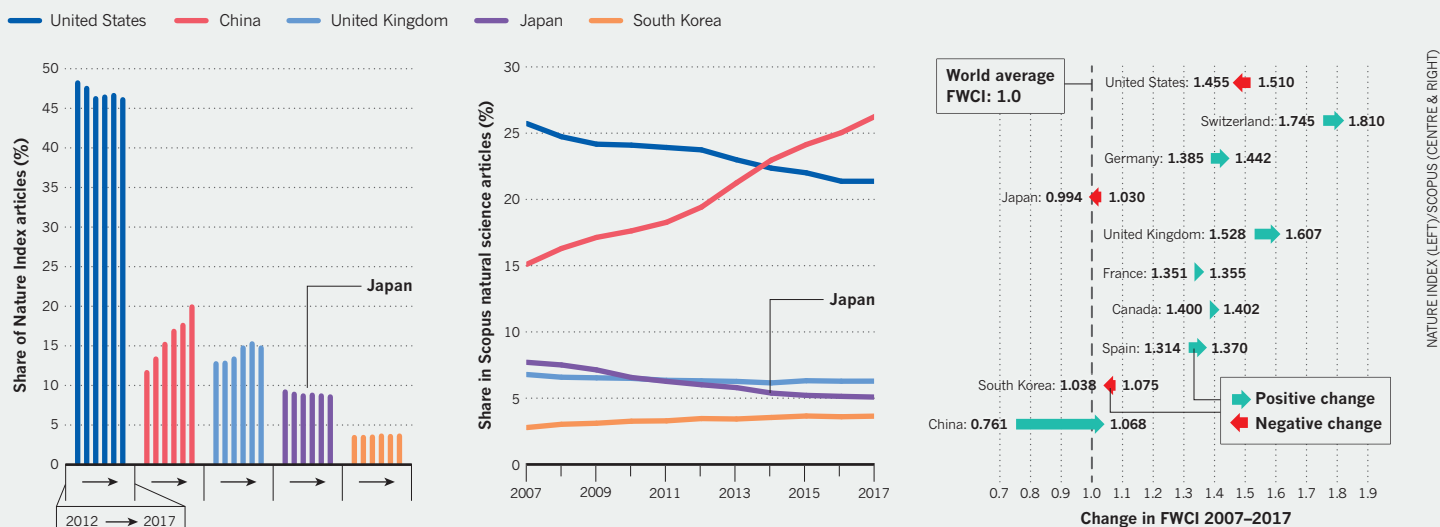
RELATIVE GAIN

While Japan's production of natural science papers continues to fall, a relatively high proportion of its output is of high quality. The country ranked 4th among the top countries in the Nature Index when assessed on its authorship of papers in the index relative to its contribution to the Scopus database, 2012–2017. Some of Japan's smaller institutions stand out for their efficiency in producing high-quality science.

DATA ANALYSIS BY AARON BALLAGH, BO WU, AND WILLEM SIJP OF THE NATURE INDEX; AND KANA TAKASAKA AND ANDERS KARLSSON AT ELSEVIER

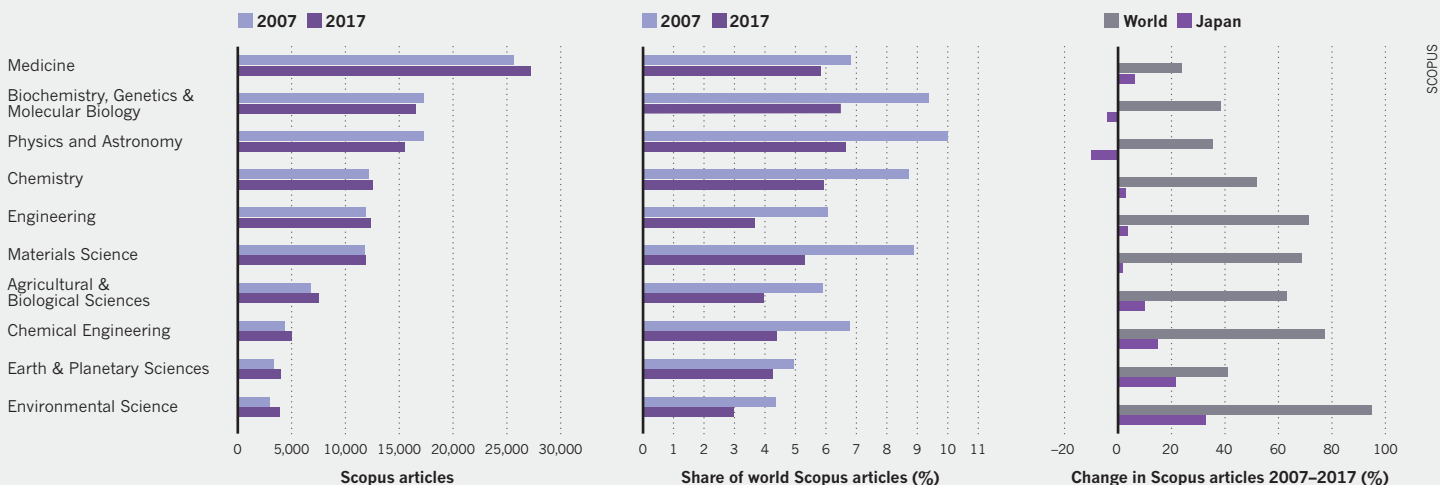
DECLINE IN GLOBAL SHARE

Japan's share of the world's high-quality research articles declined from 9.2% in 2012 to 8.6% in 2017 (left), and its share of natural science papers in Scopus also dropped over the past decade, from 7.7% in 2007 to 5.1% in 2017 (centre). The country's field-weighted citation impact (FWCI) in 2017 was lowest among the top ten countries in the Nature Index in 2017 (right).



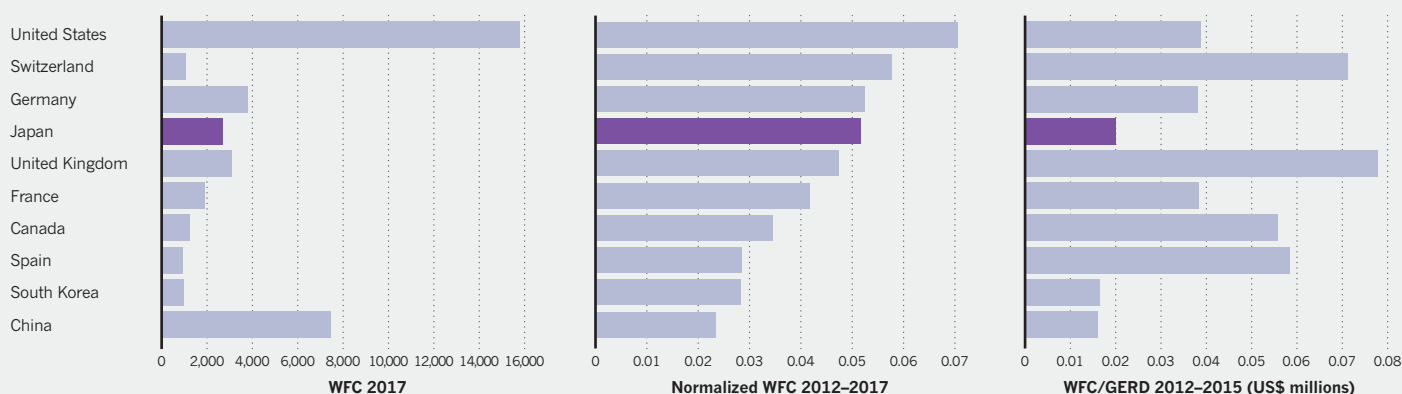
REDUCED STAKE IN EVERY SUBJECT

While Japan's article counts in the Scopus database of science articles have increased in absolute terms in the major science fields (left), its share of global output has declined in each one (centre). Shown here are Japan's top ten science subjects in the database in 2017, and their change in output since 2007 (right).



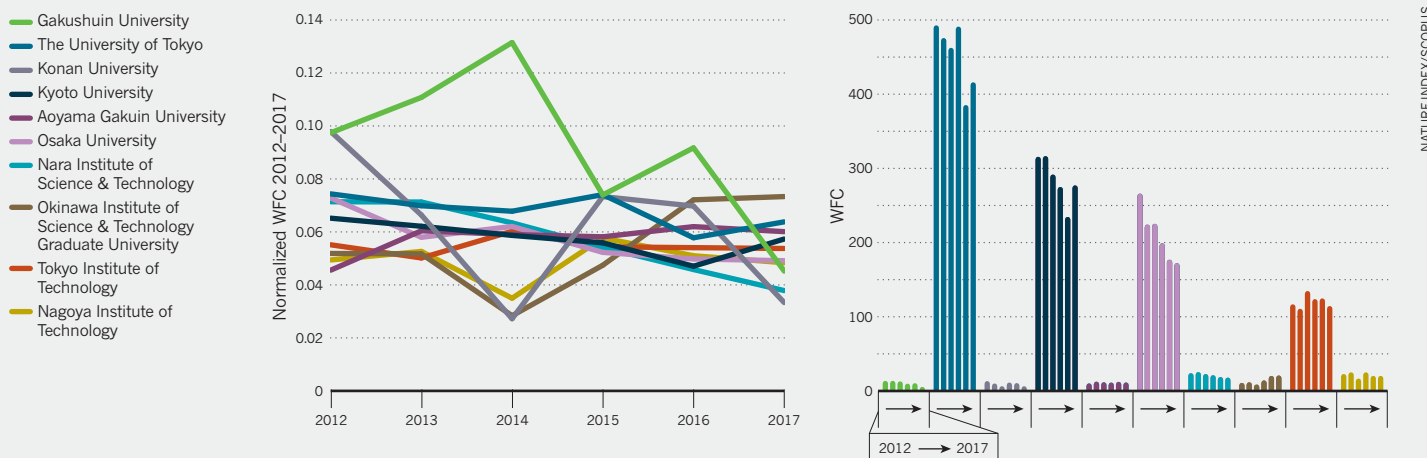
EFFICIENT PRODUCER OF QUALITY SCIENCE

Japan ranks 5th globally for its contribution to the authorship in the Nature Index, measured by weighted fractional count (WFC, left). But it jumps to 4th place among top countries in the Nature Index when assessed on the proportion of its total natural science output that is of high quality (normalized WFC 2012–2017, centre). For every individual contribution to an article in the index, Japan spent US\$50 million on research and development, double the amount spent by the United States (right).



SMALLER INSTITUTIONS SHINE

Tiny Gakushuin University cannot compete with the research behemoth The University of Tokyo for quantity of output (right). But when the playing field is levelled, to assess which Japanese institutions produced the most high-quality research relative to their total output in the natural sciences over the past six years, Gakushuin comes out on top.



A GLOBAL COMPETITOR

Top Japanese universities are on the whole more focussed on the natural sciences (NS) than some of their counterparts in the US, UK and South Korea. But article counts (AC) are declining in some. With the exception of The University of Tokyo, they produce fewer high-quality articles as a proportion of their total output in the natural sciences (NS) than top global rivals.

| Institution | Normalized WFC 2012–2017 | Nature Index WFC 2012–2017 | Scopus NS articles 2012–2017 | Nature Index AC 2012–2017 | Change in AC 2012–2017 | Scopus NS articles (%) 2012–2017 |
|---------------------------------------------------------------|--------------------------|----------------------------|------------------------------|---------------------------|------------------------|----------------------------------|
| Massachusetts Institute of Technology, US | 0.1089 | 2928.97 | 26899 | 9744 | 14.8% | 77.6% |
| Harvard University, US | 0.0767 | 4875.56 | 63562 | 15966 | -3.4% | 53.6% |
| Korea Advanced Institute of Science & Technology, South Korea | 0.0708 | 763.64 | 10792 | 1685 | 1.4% | 71.9% |
| The University of Tokyo, Japan | 0.0680 | 2699.17 | 39721 | 7829 | 0.7% | 78.1% |
| University of Oxford, UK | 0.0659 | 2260.99 | 34294 | 7594 | 17.0% | 65.0% |
| Kyoto University, Japan | 0.0577 | 1686.08 | 29212 | 4334 | -6.8% | 79.3% |
| Osaka University, Japan | 0.0574 | 1239.47 | 21579 | 3232 | -24.0% | 75.1% |
| Tokyo Institute of Technology, Japan | 0.0547 | 701.44 | 12834 | 2031 | 5.5% | 83.4% |
| Nagoya University, Japan | 0.0478 | 741.82 | 15527 | 2480 | -4.7% | 74.9% |
| Peking University, China | 0.0423 | 1698.13 | 40146 | 5725 | 105.2% | 72.9% |

NATURE INDEX/SCOPUS/OECD

NATURE INDEX/SCOPUS

NATURE INDEX/SCOPUS



COMMENT
NORIKO OSUMI

CALLING RIKEJO

A push for more Japanese women of science.

The Japanese gender gap in science, technology, engineering and mathematics (STEM) fields remains large. The Japanese Cabinet Office's 2017 White Paper on Gender Equality reports that just 10.2% of engineering researchers are women. The picture is slightly better in science (14.2%) and agriculture (21.2%). Across all fields (including in the social sciences) only 15.3% of Japanese researchers are women — about half the average proportion among OECD countries.

'Leaky pipelines' — where women drop out before establishing a sustainable career — are globally ubiquitous in STEM. In the United States, for example, a 2011 survey showed 52% of PhD students in neuroscience, my field, were women. But that drops to 44% for postdoctoral fellows and to 29% for faculty. The numbers are worse in Japan, judging by membership figures for the Japan Neuroscience Society. A 2017 survey found that only 32% of student members are women, and the proportion falls to just 20% for regular members, which includes postdocs and faculty.

UNCONSCIOUS BIAS

The under representation of women in the STEM workforce seems to originate at university. The primary and secondary school systems support each gender equally, and there is no significant difference in mathematics and science scores for Japanese students aged 12 to 15, according to the global Programme for International Student Assessment survey.

But at universities, the bias against women in STEM, dubbed *rikejo*, which roughly translates as 'science women', appears to start. Many parents and teachers believe that girls who follow the *rikejo* path will have reduced job and relationship opportunities.

The bias spills into recruitment of faculty. Male research leaders worry that hiring females may disadvantage their team in the competition for publications and funding grants. The opposite, however, may be true. An Elsevier report published in 2017 at the Gender Summit in Tokyo identified Japan as the only country where the score of scholarly output per researcher during the years 2011–2015

was higher for women than for men. Japanese women published an average 1.8 papers over the period, which was 38% more than men at 1.3 papers.

It may be a statistical anomaly due to the low numbers of women doing research. It could also be that because male researchers occupy more senior positions, more of their time is taken by administrative work rather than research. Another explanation, suggested by Reiko Kuroda, a chemist and former member of the Council for Science, Technology

and Innovation, is that women in Japanese science are exceptionally high-achieving. When I entered graduate school, my supervisor warned me that I would have to achieve twice as much as a man to establish an academic career. It's also possible the higher scholarly output of women in part reflects the occupation of faculty positions by under-productive males.

WORK-LIFE BALANCE

The gender gap cannot be bridged unless we change our work culture, especially in science and technology, where long hours in the lab are considered normal. According to OECD data, Japanese men work twice as many hours as, for comparison, French men. Because Japanese men spend less time at home, women carry a heavier load of housework and child care. A 2017 survey of staff at my university found that 64% of men spent an hour or less a day on house and child care, while 73% of women spent two hours or more.

Diversity demands the widespread adoption of a rational and reasonable work-life balance, which is also a prerequisite for a creative atmosphere that fosters innovation.

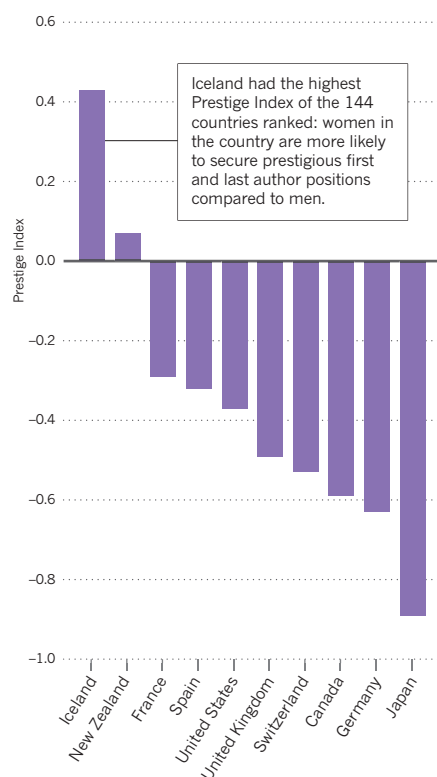
We need more women at all levels as role models — professors, deans, and presidents in academia. In 2016, universities in Japan set targets for raising the percentage of female researchers by 2021. Tohoku University, for example, plans to increase female researchers from 13% to 19%, and has begun introducing female-only faculty positions.

Tohoku also supports female researchers in various ways, including plans to increase four-fold the number of children it can accommodate in its day-care centre. Ultimately, gender equality will mean first dispelling our societal biases, and this costs universities nothing. ■

Noriko Osumi is executive director of the United Centers for Advanced Research & Translational Medicine, director of the Center for Neuroscience and a professor at the Department of Developmental Neuroscience at Tohoku University School of Medicine.

EXCLUSIVE AUTHORSHIP

Women in Japan are significantly under-represented in prestigious authorship positions in articles published between 2008 and 2016 in 54 of the 68 journals listed on the Nature Index.



SOURCE: BENDELS, M. H. K. ET AL. PLOS ONE (2018).



Daikin has enlisted the expertise of researchers at Osaka University in its quest to apply artificial intelligence to its air conditioners.

PARTNERS IN DISCOVERY

Osaka University bets on commercial-academic collaborations to survive.

BY ICHIKO FUYUNO

Osaka University's strategy for long-term sustainability in the face of shrinking government funding emphasizes a new model of collaboration with industry based on the "co-creation of values", says its president, Shojiro Nishio.

Partnerships have historically involved a university responding to a company's wish list, he explains. But companies confronting rapid societal change now seek collaboration with university experts across a wider spectrum, and in the very earliest stages of research, from discussing themes and identifying market opportunities, to training engineers and discovering drugs.

At a time when government budgets for fundamental research and staff are shrinking, Nishio claims Osaka's model for industry-academic collaborations is "the solution to strengthen innovative capacity for universities."

Osaka came top among Japanese universities for innovation in the Nature Index, ranked according to the Lens Influence Metric which measures the influence of scholarly articles on patents. It is 4th among Japanese universities when assessed on its number of partnerships with Japanese companies that have resulted in publications between 2012 and 2017. Since 2006, it has set up nearly 70 offices and laboratories on campus for joint research with companies.

Nishio took office in 2015, and made it his

mission to revitalize the university's ties with industry. In April 2017, he restructured the industry relations office to offer comprehensive support, from exploiting research ideas to patent management.

While the number of collaborative projects between universities and companies in Japan has grown from 15,544 in 2010 to 20,821 in 2015, only 3.9% of them received more than 10 million yen (US\$90,000) from companies, according to Ministry of Education, Culture, Sports, Science and Technology. Under Nishio's leadership, Osaka has signed several large contracts with companies.

In May 2016, Chugai Pharmaceutical agreed to provide one billion yen (US\$9 million) over 10 years for Osaka University's Immunology Frontier Research Center (IFReC). Government funding for the institute, under the World Premier International Research Center Initiative, ceased that year. IFReC can spend the recent influx of funds on the operation of the centre at its discretion. In return, Chugai researchers can join discussions and experiments, as well as have a first look at outcomes. IFReC has signed a similar 10-year contract to give Otsuka Pharmaceutical a second look at outcomes.

In June 2017 Daikin Industries, which is finding ways to apply artificial intelligence to its flagship air-conditioning products, agreed to invest 500 million yen (US\$4.5 million) in Osaka University, every year, for 10 years. More than 20

selected young engineers have since transferred from the company to Osaka's campus, and are discussing themes for joint research. Meanwhile, Panasonic, the electronics manufacturer, has jointly developed a curriculum with the university to teach artificial intelligence and its commercial application.

Nishio has also convinced Daikin and Biken Foundation, the university's research foundation for microbial diseases, a spin-off specializing in vaccines development, to support employment of promising young researchers for the next decade. Four researchers have already been employed under the plan and Nishio hopes to increase the number to 30.

ON A RAIL


Osaka has a history of successful ties to industry. In 1967, it launched Japan's first automated train ticket gates, jointly developed with electronics company, Omron, and the private rail services, Kintetsu and Hankyu.

Research at Osaka into interleukin 6, a protein that regulates the immune system, led to the development, along with Chugai, of tocilizumab, a drug used to treat rheumatoid arthritis, among other diseases. In 2016, researchers at Osaka and the medical devices company, Terumo, launched HeartSheet, the world's first regenerative medicine product designed to treat heart failure. Osaka's recent industrial alliances are aimed at achieving similar success. ■

BUDDHIKA WEERASINGHE/BLOOMBERG VIA GETTY

STRENGTH FROM WEAKNESS

Japan's vision of a high-tech, 'super-smart' society relies on sustained investment to arrest scientific decline.



The sky above the VERA Mizusawa Station radio telescope, following the motion of stars. Japan has a strong astronomy research output.

BY TIM HORNYAK

The Japanese government points to the country's achievements in robotics, sensor devices, supercomputers, and network and transport infrastructure as key to achieving its goal of becoming a "super-smart society", dubbed Society 5.0.

The five-year plan in effect since April 2016 is for a high-tech world in which cyberspace and real space converge, powered by the Internet of Things (IoT), big data, and artificial intelligence (AI). But analysts caution that Japan faces big challenges in enabling areas of research that have been relatively neglected. The country's share of global articles in the Scopus database has fallen in every scientific field in the past ten years. Between 2000 and 2017, the global share of papers in computer science, engineering and materials science declined by around 60% each.

Japan's investments in supporting fields such as information and communications technology have not kept pace with other advanced countries, says Hiroshi Nagano, a lecturer in science

and technology policy at the National Graduate Institute for Policy Studies (GRIPS) in Tokyo. This shows in the small contribution of computer science papers (4.4%) to Japan's overall Scopus output in 2017.

A 2016 World Economic Forum survey ranked Japan 10th in a list of countries according to their use of technology to increase competitiveness and well-being, behind states including Switzerland, Israel, Singapore, and the United States.

HUMAN RESOURCES

Building a super-smart society will also require significant investment to build skills among young people who currently do not regard science as an attractive career.

"The demand for technology skills is enormous and cannot be filled by shrinking numbers of young graduates," says Martin Schulz, senior research fellow at Fujitsu Research Institute.

A 2016 Ministry of Economy, Trade and Industry survey warns of a shortfall of 48,000 IT professionals by 2020.

The government white paper highlights the Ministry of Education, Culture, Sports, Science and Technology's efforts to nurture more professionals including AI engineers, data scientists, cybersecurity experts and entrepreneurs.

One example is the Advanced Integrated Intelligence Platform Project, a learning programme launched in 2016 with a 5.4 billion yen (US\$50 million) budget, which aims to impart the latest research techniques in AI in cooperation with institutes such as RIKEN, the Institute of Statistical Mathematics and the National Institute of Informatics.

Schulz insists that plugging the skills gap will require major corporate investment, which remains scarce, particularly among smaller companies. He points to areas such as telecommunications, smart cities and logistics in which Japan can build on strengths to realize a super-smart society.

MEDICINE AND ASTRONOMY

In 2017, the natural science subject with the largest share of the country's output in Scopus was medicine, followed by biochemistry, then physics and astronomy.

Nagayasu Toyoda, president of Suzuka University of Medical Science in Suzuka City, says that Japan's output of papers in medicine has risen because there are more doctors available to do research. University hospitals have managed to boost their income and hire, despite budget cuts. A 2016 survey by the Ministry of Health, Labour and Welfare shows that the number of medical doctors in Japan increased to 304,759 from 296,845 in 2014.

Astronomy has been helped by the construction of an enormous digital camera on the Subaru Telescope on Mauna Kea in Hawaii. In 2013, the Hyper Suprime-Cam was switched on, and by 2016, Japanese researchers had used it to identify a small galaxy orbiting the Milky Way.

Masahiko Hayashi, director general of the National Astronomical Observatory of Japan, attributes Japan's impressive astronomy output to its involvement in successful observatories such as Subaru and the Atacama Large Millimeter Array.

Hayashi also recognizes the work of the Kavli Institute for the Physics and Mathematics of the Universe. It was established by the government as one of nine centres of excellence for interdisciplinary research under the World Premier International Research Center Initiative, which in 2018 will fund 11 centres at a cost of 60 billion yen (\$US560 million). ■

NATIONAL ASTRONOMICAL OBSERVATORY OF JAPAN

PILLARS OF A SMART SOCIETY

Across the scientific fields, Japanese researchers are bringing together the virtual and real worlds.

BY SMRITI MALLAPATY

Society is ready, says Japan, to enter its next stage of evolution, in which the real world is more intimately connected to the virtual world. In Society 5.0, as laid out in Japan's five-year plan, perennial problems of inequality, incohesion and estrangement will be left behind with the help of emerging technologies. People of all ages and ability will gain opportunities for enhancement, whether from superhuman physical strength, exclusive medical insights, untethered mobility, or informed decision-making.

For Japan, a super-smart society is not just an aspiration, but a necessity. Its proportion of people older than 65 is expected to reach nearly 40% by 2065, which will stretch the country's workforce beyond capacity. The work of the researchers featured here points to what Society 5.0 might look like.

KANAKO HARADA

Bionic builder

Kanako Harada is building a mannequin that mimics the inner workings of the human body. The 'bionic humanoid' offers a substitute for the use of cadavers and animal models in research and surgical practice.

The 41-year-old engineer at The University of Tokyo is one of 16 project managers appointed under a Japanese government initiative inspired by the United States Defense Advanced Research Project Agency (DARPA). ImPACT (Impulsing Paradigm Change through Disruptive Technologies) was launched in 2015 with a budget of 550 billion yen (US\$5 billion) to back transformational ideas. Harada now leads a team of 150 researchers from Japanese and French institutions, companies and hospitals.

Other ImPACT projects include a high-speed smartphone sensor that can detect hazardous molecules, bacteria, viruses, and air pollutants, and the ultimate efficient computer that only needs to be charged monthly.

The idea that robotic technology could mean better surgical outcomes has excited Harada since she first saw the da Vinci robotic surgical arm, designed by the Californian company, Intuitive Surgical. "It was like a dream come true," she says.

Surgeons have been using da Vinci for more than a decade to conduct minimally intrusive operations. Harada's team is developing similar robotic arms for sectioning and suturing the delicate tissue of the brain and eye. "No robots exist yet for this kind of surgery," says Harada.

Her two-armed system can precisely grasp a range of surgical instruments, including a 3.5-millimetre tip that slips through the nostrils like a snake to conduct microsurgeries. The goal of the three-year project, says Harada, is for her robots to achieve a level of accuracy and reach that humans cannot.

A prototype of a bionic eye developed by Harada's team has an inner membrane, several micrometres thick, which can be peeled using tiny metal graspers, the way surgeons would operate on an eye to improve vision. Ophthalmologists can also train using the model.

Harada hopes to see her surgical systems in operating theatres within the next decade.

NAOKI SUGANUMA

Test driver

Naoki Suganuma's laboratory has clocked more than 10,000 kilometres of autonomous driving — a distance roughly four times the length of Japan — in less than two years. Suganuma, a 42-year-old engineer at Kanazawa University,

led the first Japanese tests of self-driving cars on public roads in February 2015. A Toyota Prius, fitted with Suganuma's satellite, radio, laser, and optical navigation system drove along busy urban streets and mountain roads, by day and by night. It was the first of many tests aimed at designing autonomous vehicles that could be used as a form of public transport.

"In 10 or 20 years, we could face a shortage of drivers," says Suganuma, referring to Japan's ageing population. The country became the world's first super-aged society in 2006, with more than one in five people above the age of 65. By 2065, that number is expected to rise to two in five.

In the two decades since Suganuma started working on autonomous driving technologies, his laboratory has developed expertise in every aspect of the autonomous driving system, from the creation of digital maps, to perception, planning and control systems, and human-machine interfaces. It has also collaborated with many companies, including Toyota, Denso Corporation, and Sony.

For Suganuma, the most pressing problem is snow. In a white landscape autonomous





Kanako Harada and the bionic humanoid developed for surgeons to practise operative procedures.



A wheelchair developed at the University of Tsukuba can transfer users into an upright position.

vehicles struggle to detect lines on roads, and distinguish roads from footpaths. His team is developing techniques to improve the positioning of vehicles along lanes in rainy and snowy conditions.

Thanks to the strong support of government for developing the technology, notably from the prime minister, Shinzo Abe, “Japan is the easiest country in the world to conduct public road tests of autonomous vehicles,” says Suganuma. A strong automobile manufacturing base also gives it an edge in developing the complex systems needed to make effective self-driving cars, and compete with the United States and Germany.

Using data gathered from cars, says Suganuma, researchers in Japan could develop even more sophisticated artificial intelligence technologies. That is, he says, if companies agree to share their proprietary information for big-data mining. So far, car-makers have been reluctant to release their sensor data.

KENJI SUZUKI

Social enabler

Kenji Suzuki tries to see eye to eye with his research subjects, sometimes even literally. The physicist heads the Artificial Intelligence Laboratory at the University of Tsukuba. His multidisciplinary team, which includes computer scientists and electrical and biomedical engineers, as well as psychologists and philosophers, has designed a wheelchair that can transform into a roving platform. The mobility vehicle will allow people with lower-limb disabilities to meet people at eye height, seated or standing. “We don’t design robots to replace people, we design robots to empower them,” says Suzuki.

If all goes well, the vehicle will appear at the 2020 Paralympic Games in Tokyo. It is just one of the technologies that Suzuki’s team is working on to enhance social interactions among people of diverse ages and physical ability. Other innovations include a technique for detecting facial expressions via the electrical signals generated by muscles, without relying on more overt visual cues that might not be as apparent in people with visual impairments or facial paralysis.

His team has also worked with hospitals to design a swallowing monitor. The artificially intelligent, horseshoe-shaped device wraps around the back of the neck to process sounds produced while swallowing. It can alert doctors, via a smartphone app, to any gulping errors, to head off the risk of pneumonia when food enters the lungs.

In 2015, Suzuki, 43, received funding from the Japan Science and Technology Agency for a start-up to bring the device to market. The company is expected to launch in April 2018.

With the Japanese government funnelling

resources into Society 5.0, which prioritizes applied science and industry collaboration, Suzuki worries that young researchers under pressure to demonstrate value may not have the chance to cultivate ideas from basic research. “I was lucky,” says Suzuki. “I spent 10 years developing these fundamental technologies, and now they are ready to transfer to society.”

NAGANO HIROFUMI

Blockchain developer

Cryptocurrency headlines can be terrifying for investors, from stories of people being locked out of millions of dollars’ worth of bitcoins to hack attacks costing digital coin exchanges assets worth billions of yen. But researchers at Hitachi are developing technologies that could mollify investors’ fears.

Blockchain, the technology that underpins cryptocurrencies, offers a way to execute financial transactions without an intermediary, such as a bank, at a significant saving of cost and time. In a blockchain system, transactions are accessed via private and public ‘keys’, made up of very long numbers. The problem, as the headlines attest, is that these keys are kept on personal computers or servers where they can be lost or leaked. Hitachi’s system generates a digital signature using, for example, the unique vein pattern on the surface of a finger. Investors can access funds with the biometric key, which is harder for hackers to intercept. The technology, still in development, will “realize a secure and convenient blockchain system,” says Nagano Hirofumi, a 44-year-old researcher at Hitachi in Tokyo.

Hitachi, the eighth-ranked corporate institute in the Nature Index, when assessed on its high-quality research relative to its total output in the natural sciences, began researching blockchain technology in 2015. Hirofumi and colleagues saw that blockchain’s disruptive potential extends beyond the financial sector, even connecting industries using so-called smart contracts. “We can realize reliable and smart social infrastructure using blockchain,” he says. Self-executing digital intermediaries could be used to set conditions in transactions; they might, for example, release funds deposited by a buyer only after receiving items from a supplier.

The challenge will be to design appropriate regulation. Information on blockchain platforms, including private data, is globally accessible. But many countries prohibit the export of such data, even if they are encrypted, says Hirofumi.

The Japanese government is discussing the possibility of a regulatory ‘sandbox’, he says. It would be a safe space in which to trial a framework, similar to those launched in the United Kingdom and Singapore, to encourage innovation in a controlled legislative environment. ■

FACING DOWN DISASTER

An institute's work on warning and mitigation systems for catastrophic events is a national priority.



Emergency workers in Kesennuma, northeastern Japan, in the wake of the earthquake and tsunami of March 2011. Close attention is paid to research into forecasting natural disasters.

BY TIM HORNYAK

Volcanic eruptions, typhoons, floods, mudslides, fires and earthquakes are commonplace in Japan. Disaster research is a national priority for good reason.

The National Research Institute for Earth Science and Disaster Resilience (NIED) is Japan's best performing government institution when assessed on its authorship of papers in the index relative to its total output in the Scopus database of natural science publications. Between 2012 and 2017, it contributed to a weighted fractional count of 22 in the index, from 304 Scopus papers in the natural sciences.

NIED was established in 1963, a few years after the Ise Bay Typhoon that killed more than 5,000 people. In 2015, it specified its mission to pursue research that will "increase the level of science and technology for disaster risk reduction" and build disaster resilience. NIED promotes research for all hazards in all phases, says its president, Haruo Hayashi. Researchers in the natural sciences, engineering and social sciences are studying better ways to predict, prevent and respond to calamities of all types, he says.

With headquarters in Tsukuba, northeast of Tokyo, NIED has a staff of about 250, and an annual budget of 9.2 billion yen (US\$81.6 million), down from 11 billion yen in 2008. Its labs across the country include the Shinjo Cryospheric Environment Laboratory in Yamagata Prefecture, the Snow and Ice Research Center in Niigata and in Miki City, and the Hyogo Earthquake Engineering Research Center, which hosts the world's largest earthquake simulator, or 'shaking table', known as E-Defense.

NIED maintains a nationwide seismographic network of more than 2,000 stations for quake monitoring. It has recently installed a 5,700 kilometre fibre-optic network along the ocean floor, connected to 150 observation stations, intended to predict tsunamis off Japan's eastern coast.

NIED researchers' publications reveal the diversity of its work. A 2015 *Nature* letter by NIED geophysicist, Futoshi Yamashita, and colleagues looked at the relationship between earthquakes and the frictional properties of rocks. Experiments on samples suggested that natural faults weaken an order of magnitude faster than had been previously estimated.

In October 2017, an issue of the *Journal of*

Disaster Research was dedicated to NIED's work. The papers covered topics such as under-sea seismic velocity around Japan, the use of virtual reality to simulate damage from earthquakes, and how information was gathered and shared in the aftermath of the quakes that hit Japan's Kumamoto Prefecture in 2016.

In a September 2016 evaluation, Japan's science ministry praised NIED for its insights and information-sharing following the Kumamoto disaster, which killed dozens and injured thousands. It also recognized the 1.54 million hours of testing conducted using the E-Defense system. E-Defense has become an essential tool for ensuring the safety of structural designs, says Masayoshi Nakashima, former director of the operation currently at Kyoto University.

But, NIED has its critics too. Seismologist, Robert Geller, professor emeritus at the University of Tokyo, has been critical of Japanese efforts to forecast earthquakes "because of the dodgy nature of the underlying science." He says NIED's network of observatories provides essential data for seismologists but the network's funding "should be continued based on its own merits, rather than the pretence that it's needed to forecast earthquakes". ■

NEWSCOM/ALAMY

SHORT-TERM GENERATION

With permanent academic jobs scarce, young researchers on temporary contracts are nervous about the future. The government wants industry to utilize their talents.

BY SMRITI MALLAPATY

For the past three years, Hirotaka Kawashima has been worried about his fate. The 37-year-old scientometrics researcher from Tokyo is among the thousands of early-career researchers in Japan on a short-term work contract. He may soon be out of a job.

"I am very worried," says Kawashima, who has a doctoral degree in cognitive neuroscience and is employed as a research fellow at the National Institute of Science and Technology Policy (NISTEP) until March 2018. "I have no future positions," he says. "Most of my researcher friends have fixed-term positions," says Kawashima.

Kawashima and his colleague, Yasuhiro Yamashita, who is also on a fixed-term contract, recently analysed the academic employment database JREC-IN Portal, which is maintained by the Japan Science and Technology Agency. They found that 71% of the job postings for assistant professors, published between 2012 and 2015, were for limited terms.

With few prospects for landing a permanent job in academia or national research institutes, recent graduates and postdoctoral researchers in Japan are being forced to abandon their scientific ambitions.

The talent drain will make it more difficult to arrest the decline in Japanese science. A separate NISTEP study found that between 2004 and 2012, junior scientists contributed to almost 70% of the papers published by the top 44 most productive Japanese universities in the Web of Science database.

"Early-career researchers are important for conceiving and testing new ideas and introducing disruptive innovation. We need to create better policies that support them," says Yuko Harayama, an executive member of the prime minister's science advisory body, the Council for Science, Technology and Innovation.

In Japan, unlike many other research-intensive countries, PhD students largely have to cover their own school fees, pension and living expenses. Around half of postdocs are lucky enough to score a permanent position, but the



A researcher at the Control System Engineering Laboratory, at Hiroshima University, which is a big supporter of linking graduates with industry.

best the rest can hope for after graduation is a five-year fixed contract, the longest that are generally permissible under the Labor Standards Act.

In 2016, the government set a target to increase the number of university researchers under 40 by 10% by 2020 from a starting point of 43,763 researchers in 2013.

GOING PRIVATE

The Japanese government has introduced several initiatives over the past decade to increase the number of permanent jobs available to young researchers. But these actions have yielded few measurable gains, says Harayama. Now, the government is making overtures to

the private sector with a view to taking graduates outside academia entirely. "We really want to invite companies to be a part of this story," says Harayama.

Toshiyuki Misu, a physicist who heads the Global Career Design Center at Hiroshima University, is facilitating this transformation. Before moving to Hiroshima in 2014, Misu spent nine years at NISTEP studying the Japanese graduate employment landscape.

He and his colleagues analysed the career paths of a quarter of the country's postdoctoral population, finding that very few doctorate graduates ended up in industry.

Of the more than 3,800 researchers surveyed, only 68 joined the private research and development workforce. Three-quarters continued to work as postdoctoral researchers in academia. A 2012 survey of the entire Japanese postdoctoral population proved consistent: just 106 of the almost 13,000 researchers whose careers could be traced took up R&D positions in a company.

Since Misu's study, the government has introduced several initiatives to encourage universities to form connections with companies as a conduit for science careers for younger generations.

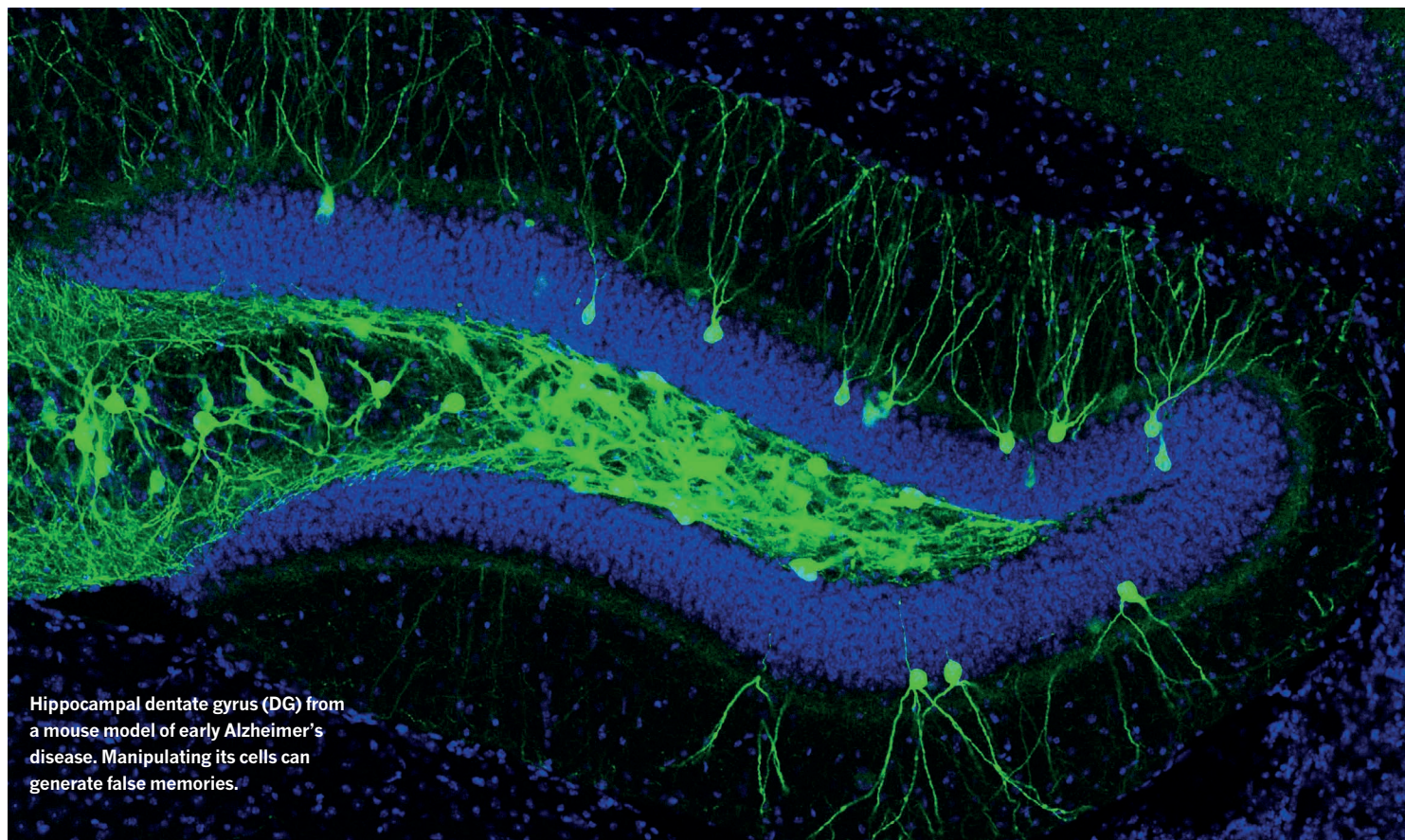
Hiroshima has supported training on leadership and communication skills and internships at partnering companies for nearly 100 graduate students. The programmes are gradually convincing students, faculty, and companies of the merits of bringing

PhD graduates into the corporate sector, says Misu. "We should inform industry that it needs these highly skilled researchers for innovation."

More movement is afoot. In 2016, the government launched the Leading Initiative for Excellent Young Researchers to match 100 young scientists with tenure-track programmes in academia, government or industry every year, with the option for joint appointments.

And in 2017, it started a five-year programme to encourage young researchers to establish start-ups. "Young people are highly motivated to contribute to society and global challenges," says Harayama. "We have to make sure that private companies understand that it is to their advantage to include them." ■

TAMOTSU KASHIWAGI



Hippocampal dentate gyrus (DG) from a mouse model of early Alzheimer's disease. Manipulating its cells can generate false memories.

ACCESS GRANTED

International collaborations are rising, but not enough to put Japan on the international researcher circuit.

BY DANIEL HURST

The Japanese government recognizes the relative isolation of Japanese science as a problem.

"Strategic international development combined with science and technology diplomacy is indispensable for Japan," asserts the 2016 White Paper on Science and Technology, a blueprint for Japan's goal of becoming a 'super-smart' society.

But the call is not new, as disaster management researcher Rajib Shaw points out. He has witnessed the country's efforts to deepen its global connections over nearly 25 years, since coming to Japan as a student. In most cases, new schemes to promote collaboration are simply repackaged versions of old schemes, he says: "It's like old wine in a new bottle."

Shaw has spent 14 years working in academic institutions in Japan and is now professor at Keio University, and chair of the UN Office for Disaster Risk Reduction's Science Technology Advisory Group. He leads a component of a

joint Japanese-Bangladeshi project, looking at flood risk in Bangladesh.

Mindsets need to change, says Shaw. Not enough Japanese researchers are eager to publish their papers in English and present to international conferences, he says. This reduces their impact.

PAIRING UP

Nature Index data show the government has made some progress against these goals, although the results are mixed. Only about half of the papers published by Japanese researchers have an international co-author, compared with about three-quarters of papers in the United Kingdom and Germany.

However, the number of international partnerships involving Japanese researchers has risen from 20,106 in 2012 to 33,439 in 2017 — an increase of more than 65%. The United States remains Japan's most frequent collaborator in producing high-quality articles across four broad science categories, although the pair's

total collaboration score declined 4.7% over the same period.

In 2013, for instance, researchers from the RIKEN-Massachusetts Institute of Technology (MIT) Center for Neural Circuit Genetics reported in the journal *Science* on how they could manipulate cells in the mouse hippocampus to generate false memories. More recently, they have published papers on memory in *Nature* that offer potential clues about depression and Alzheimer's disease.

The centre, headed by Nobel laureate, Susumu Tonegawa, brings together researchers from the RIKEN Brain Science Institute in Saitama and the Picower Institute for Learning and Memory at MIT. Over the past six years, RIKEN has engaged in 1,868 partnerships with international institutions, making it Japan's second most collaborative institution after The University of Tokyo.

Nature Index data show Japan's collaboration score with China surged by 41%, ranking it as Japan's third favourite collaborator after

Germany. Chinese and Japanese researchers have collaborated on projects looking, for example, at cancer and hepatitis B and C risk factors and control measures. The top-ranked areas of collaboration between the two countries in the index were in the physical sciences and chemistry fields.

BRAIN GAIN

One way of raising Japan's profile, according to the Science and Technology Basic Plan, is to "bolster global brain circulation" — which means encouraging Japanese researchers to conduct research abroad and international researchers to do the same in Japan. As part of these efforts, the government-funded Japan Society for the Promotion of Science helps organize exchanges for young researchers.

The government's World Premier International Research Center Initiative (WPI) also aims to build "globally visible" research centres that attract leading researchers from abroad. Mathematician, Motoko Kotani is a director of one of these centres, the Advanced Institute for Materials Research (AIMR) at Tohoku University.

She says 40% of its members are non-Japanese. By comparison, the average share of foreign faculty members in Japanese national universities is 4%, according to figures

compiled by the Japan Association of National Universities.

While individual Japanese researchers have contributed to global academic communities, institutions have a limited international presence, says Kotani.

Much effort has gone into supporting premier universities to work internationally, says Akiyoshi Yonezawa, the director of the Office of Institutional Research at Tohoku University. But he worries a side-effect will be a greater divide between the top- and second-tier universities.

"WE ARE STILL FAR TOO ISOLATED."

"The top universities may be in a process of changing, but at the second-tier universities performance is going down quite rapidly," he says. He points to analysis by Professor Nagayasu Toyoda showing a growing disparity in the number of clinical medicine papers published by the higher and lower-ranked university departments.

"This is very dangerous because research performance cannot be sustained only by the top universities. We need to have a larger pool

of talents." Yonezawa's concerns are not yet reflected in the Nature Index, as output from the group of 11 top research universities has declined at a similar pace to all other academic institutions between 2012 and 2017.

INTERCEPTING TRAFFIC

The most obvious obstacle to increasing Japan's international collaboration is the language barrier. Most of the older generation have "a very serious problem with English, especially in humanities and social sciences," says Yonezawa.

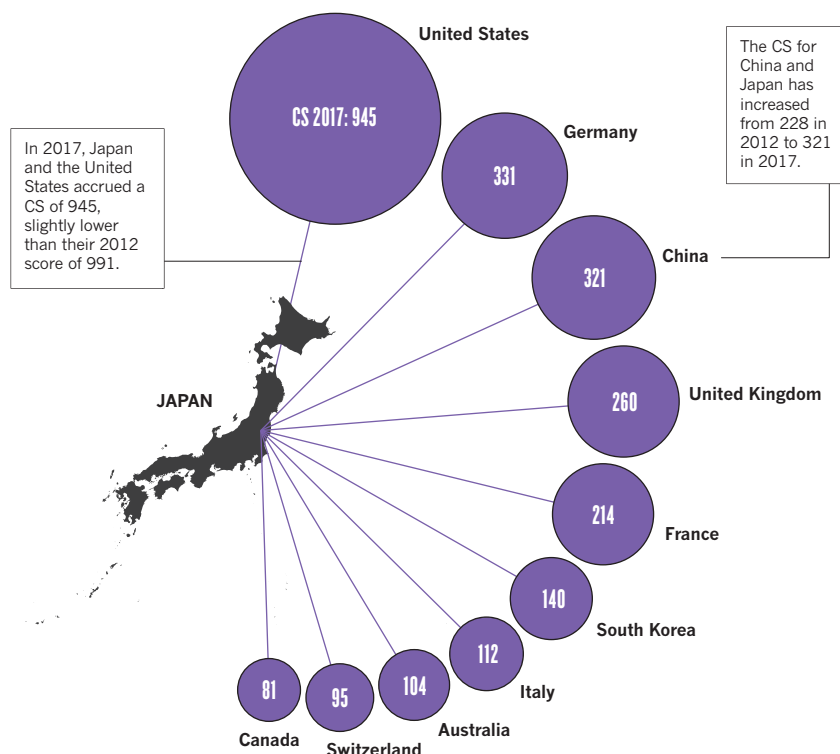
While younger generations are gradually gaining better English skills, "compared with other east Asian countries or south-east Asian countries, we are still too far isolated," he says.

Shaw of Keio says the language barrier was a "major issue" when he first came to Japan. He found that students and some senior researchers hardly read any English literature. The situation is gradually changing, although he says he has seen several researchers giving international presentations about global collaborative projects with slides in Japanese.

Kotani says the AIMR has sought to accommodate international members by adopting English as the official language, but adds: "most Japanese research grants can only be applied for in Japanese, and there are few bilateral or international joint research grants." ■

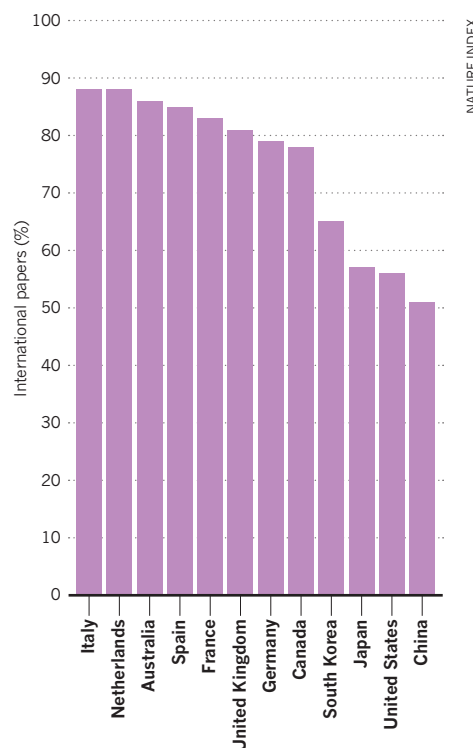
JAPAN'S TOP COLLABORATORS

Over the past six years, Japan has cultivated strong scientific links with the United States, though the number of papers the two countries have co-authored has declined slightly. The collaboration score (CS) is derived by summing the contribution of authors to articles in the index in both countries.



INTERNATIONAL COOPERATION

Slightly more than half of papers authored by researchers in Japan in 2017 had a co-author outside the country. This proportion is lower than the leading countries in the Nature Index except China and the United States.



NOBLE HALLS OF DISCOVERY

Gakushuin University is the top-ranking institution when assessed on the quality of its research in the natural sciences.

BY TIM HORNYAK

Gakushuin University takes great pride in its pedigree. Established in Kyoto in 1847 to educate Japan's nobility, its forerunner, Gakushuin School, was relocated to Tokyo 30 years later, but only opened its doors to the general public as a university after World War II.

The university's alumni include artists such as Yoko Ono, animé legend Hayao Miyazaki, and numerous politicians, as well as Japan's 84-year-old emperor, Akihito. The scientifically minded monarch has a deep interest in fish and has written extensively about the classification of gobies.

Gakushuin ranks first among all Japanese universities in the Nature Index for its

high-quality research output relative to its total number of natural science articles in the Scopus database. To achieve this, between 2012 and 2017, the university accrued a weighted fractional count of 42 for its authorship of papers in the index, out of 452 natural science papers tracked by Scopus.

The university's biggest strength is in chemistry, with some notable contributions to the physical and life sciences as well. In 2013, biologist Kiyokazu Agata and colleagues elucidated the signaling pathways involved in the flatworm's ability to regenerate tissue — an entire organism can grow out of the fragment of a tail. In 2017, biophysicists, Daisuke Nakane and Takayuki Nishizaka discovered that the hair-like appendages that blue-green bacteria use to propel themselves protrude in response to blue light.

"I WOULD NEVER BE ABLE TO MAINTAIN THIS STYLE IF I HAD MOVED TO ONE OF THE MAJOR NATIONAL UNIVERSITIES."

And in 2016, theoretical physicist Hal Tasaki co-authored a paper in *Physical Review Letters* on the fundamental relationship between power and efficiency of heat engines. A paper Tasaki published in the same journal in 1998 on the derivation of statistical mechanics based entirely on quantum mechanics went unnoticed for nearly a decade. It has since been recognized as a seminal work in the field.

Tasaki attributes Gakushuin's success to an environment that fosters curiosity for curiosity's sake. In the 30 years that he has been at the physics department, Tasaki says that he never felt pressured to publish lots of papers in high-impact journals. "Instead, I was told by older professors that we must spend a long time to accomplish something important." He would sometimes go years without publishing his work. "I would never be able to maintain this style if I had moved to one of the major national universities," he says. ■



Emperor Akihito in 1982, then Crown Prince. Akihito was a taxonomic researcher of gobioid fish, and is among many alumni of note from the prestigious Gakushuin University.

NEWSCOM / ALAMY

A GUIDE TO THE NATURE INDEX

A description of the terminology and methodology used in this supplement, and a guide to the functionality available free online at natureindex.com

The Nature Index is a database of author affiliations and institutional relationships. The index tracks contributions to articles published in a group of highly selective science journals, chosen by an independent group of active researchers.

The Nature Index provides absolute and fractional counts of publication productivity at the institutional and national level and, as such, is one indicator of global high-quality research output. Data in the Nature Index are updated regularly, with the most recent 12 months made available under a Creative Commons licence at natureindex.com. The database is compiled by Springer Nature. The list of journals tracked by the Nature Index will be extended in 2018.

NATURE INDEX METRICS

There are several measures provided by the Nature Index to track affiliation data. The simplest is the article count (AC). A country or institution is given an AC of 1 for each article that has at least one author from that country or institution. This is the case regardless of the number of authors an article has, and it means that the same article can contribute to the AC of multiple countries or institutions.

To get a sense of a country's or institution's contribution to an article, and to ensure they are not counted more than once, the Nature Index uses the fractional count (FC), which takes into account the share of authorship on each article.

The total FC available per paper is 1, which is shared among all authors under the assumption that each contributed equally. For instance, a paper with 10 authors means that each author receives an FC of 0.1. For authors who have joint affiliations, the individual FC is then split equally between each affiliation.

Another measure used is the weighted fractional count (WFC), which applies a weighting to the FC to adjust for the over-representation of papers in astronomy and astrophysics. Four journals tracked by the index in these disciplines publish approximately five times the number of papers relative to other disciplines. Therefore, although the data for astronomy and astrophysics are compiled in the same way as for other disciplines, articles from these journals are assigned one-fifth the weight of other articles (the FC is multiplied by 0.2 to derive the WFC).

The total FC or WFC for an institution is calculated by summing the FC or WFC for individual authors. The process is similar for

natureindex.com users can search for specific institutions or countries and generate their own reports, ordered by article count (AC), fractional count (FC) or weighted fractional count (WFC).

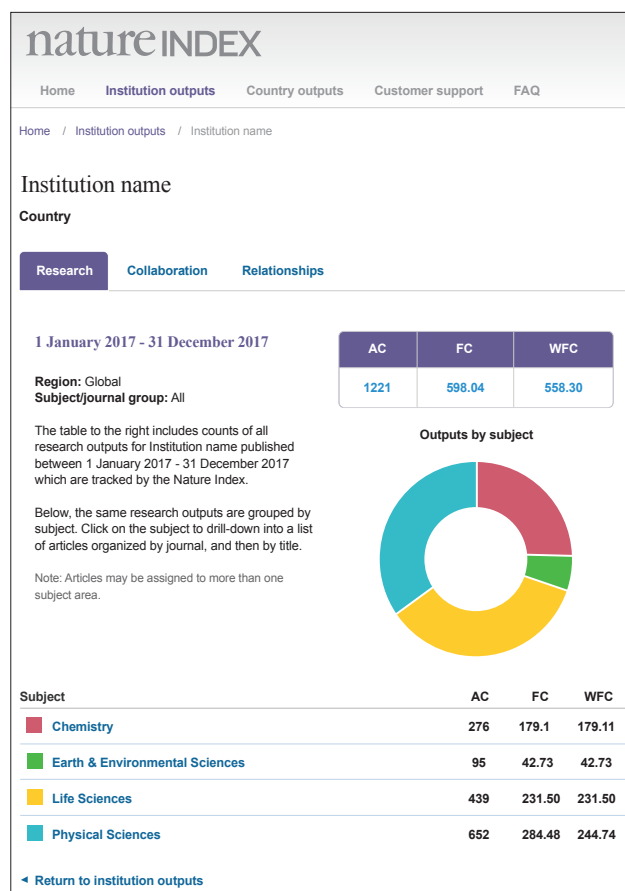
Each query will return a profile page that lists the country or institution's recent outputs, from which it is possible to drill down for more information. Articles can be displayed by journal, and then by article. Research outputs are organized by subject area. The pages list the institution or country's top collaborators, as well as its relationship with other organizations. Registering allows users to track an institutions' performance over time, create their own indexes and export table data.

countries, although complicated by the fact that some institutions have overseas labs that will be counted towards host country totals. The fourth metric is bilateral collaboration score (CS). A bilateral collaboration can be between any two institutions or countries co-authoring at least one paper in the journals tracked by the Nature Index. CS is derived by summing the FCs from papers with authors from both institutions. If institution A has co-authored a paper with another institution, then the collaboration score between A and B is the sum of the FC for A+B.

Every effort is made to count affiliations consistently, with a background of reasonable assumptions. For more on how the affiliation information is processed, visit natureindex.com.

NATUREINDEX.COM

A global indicator of high-quality research



THE SUPPLEMENT

Nature Index 2018 Japan is based on data from natureindex.com, covering articles published during six years from 1 January 2012 to 31 December 2017. The tables rank Japan's top 100 institutions by normalized WFC, an indicator of an institution's high-quality research output as a proportion of total output in the natural sciences. This is derived by dividing an institution's cumulative WFC in the Nature Index 2012–2017 by its total number of natural science articles in Scopus 2012–2017. Also listed are the institution's total number of articles in the Nature Index and the percentage of its total articles in Scopus that are in natural sciences. Separate tables rank institutions in academic, corporate and other sectors, as well as in specific subject areas. ■